**MINISTERE DE L'AGRICULTURE**
ECOLE NATIONALE SUPERIEURE AGRONOMIQUE DE MONTPELLIER

**T H E S E**

pour obtenir le DIPLOME DE DOCTORAT

*Discipline* : Biologie des populations et ecologie
*Formation Doctorale* : Biologie de l'évolution et écologie
*Ecole Doctorale* : Biologie des Systèmes Intégrés Agronomie Environnement

présentée et soutenue publiquement
par

**Marie-Pierre Chapuis**

le 20 juin 2006

_____

# Génétique des populations d'un insecte pullulant, le criquet migrateur, *Locusta migratoria*

_____

JURY

M^r^ **Kreiter** Serge, Enseignant Chercheur, Agro-Montpellier, Président
M^r^ **Michalakis** Yannis, Directeur de recherche, CNRS, Directeur de Thèse
M^r^ **Goudet** Jérôme, Professeur, Université de Lausanne, Rapporteur
M^r^ **Simon** Jean-Christophe, Directeur de recherche, INRA, Rapporteur
M^me^ **Samadi** Sarah, Chargé de recherche, IRD, Examinateur
M^r^ **De Loof** Arnold, Professeur, Université Catholique de Louvain, Examinateur

Ma planète thèse a de la chance : trois étoiles ont veillées sur elle. *Merci à Arnaud, Yannis, Michel.*

De précieux satellites ont aussi traversé son univers : membres du jury et des comités de thèse, et accros de la pipette, des stats, du criquet en boîte, ou du filet. *Merci à Arnold De Loof, Gérôme Goudet, Serge Kreiter, Sarah Samadi, Jean-Christophe Simon, Thierry Boulinier, Patrice David, Nicolas Galtier, Yves Gillon. Merci à Anne, Laurent, Sylvain, Antoine, Arnaud, Alex, S. Arsol, Mr. Badaruddin, L. Baoping, D. Brown, N. Cliquennois, O. Deveci, A. El-Heneidy, P. Fontana, A. Guendouz-Benrima, O. Hasbullah, Y. Hidayat, D. Hunter, Mr. Kadiran, I. Kamal, Mr. Karsono, A. Khan, P. Meunier, O. N. Naumovich, D. Llucià Pomares, F. Ramli, T. Rachadi, G. Reynaud, X. Shengquan, S. Soumavro, Mr. Sumarli, F. Tami, S. Tanaka, S. Yskak, T. Yulinati, S. Wangpeng, Mr. Widji, Mr. Winarto, D.-X. Zhang, G. Zhiyong.*

Si je ne me suis pas égarée, c'est aussi grâce à mon petit noyau familial à la gravité si astronomique. Bien sûr il y a eu les autres astres vagabonds qui quelque soit leur voyage ont donné tant de souffle au mien. Mon caillou n'a plus qu'à tourner sur lui-même, il devrait bientôt se retrouver la tête en bas – c'est définitivement une année qui donne à chavirer. *Merci à mes parents, mon frère, ma sœur, Aymeric, Sandrine, Claudie, Claire, Céline, Fred, Thomas, Sophie, Emilie, Guillaume, Eric, Olive, Juju, Ozge, Armelle, Joëlle, Karine, Lolo et Eleni.*

# Sommaire

# Liste des figures

# Introduction

# 1. LES PULLULATIONS D'INSECTES

Les insectes devancent tous les autres animaux en tant que ravageurs des plantes cultivées et forestières. Pour certains de ces insectes, l'impact écologique et économique majeur est lié à leur introduction et leur dispersion par l'homme. On parle alors d'espèces exogènes. Parmi les espèces exogènes les plus dévastatrices, on peut citer l'anthonome du cotonnier (*Anthonomus grandis*), le doryphore (*Leptinotarsa decemlineata*), et la mouche méditerranéenne des fruits (*Ceratitis capitata*). A ces espèces exogènes s'opposent les espèces natives remarquables par des fluctuations extrêmes des densités de leurs populations en l'absence d'action anthropique directe apparente. Ce sont ces espèces natives pullulantes qui nous intéressent plus particulièrement ici. Les changements de tailles de populations de ces espèces natives, soudains et en apparence inexplicables, ont historiquement autant fasciné qu'affecté l'humanité (Turchin 2003). La Bible décrit ainsi la menace exercée par les nuages de criquets sur l'homme. Dès le début du XX$^{ème}$ siècle, les pullulations du bombyx disparate (*Lymantria dispar*) et des scolytes (Scolytidae : *e.g.*, le scarabée des pins *Dendroctonus frontalis*) inquiétèrent les forestiers d'Europe, d'Afrique du Nord et d'Amérique du Nord et Centrale du fait des dégâts occasionnés lors de défoliation de forêts entières.

## 1.1. Définition et importance socio-économique

D'un point de vue écologique, une pullulation correspond à une augmentation explosive, d'une magnitude de plusieurs ordres, des densités de populations d'une espèce sur une période brève (Berryman 1987). Les populations d'espèces pullulantes se caractérisent par une récurrence de périodes de pullulation, aux densités de populations élevées, et de périodes de rémission, aux densités de populations faibles. L'intervalle entre les périodes de pullulation et les périodes de rémission est généralement irrégulier (Itô 1968) et supérieur à une année (*e.g.* papillons forestiers : périodicité de 8 à 11 ans, Myers 1988 ; locustes : jusqu'à une vingtaine d'années, *e.g.* Randriatmanantsoa 1998). La croissance et l'effondrement rapides des populations sont souvent associés à des temps de génération courts (*e.g.*, *D. frontalis* : 6 générations par an ; *L. migratoria* : 5 générations par an en région tropicale ; Lecoq 1975) et à de fortes

fécondités (Mason 1987). Le caractère explosif et pluriannuel des fluctuations permet de discriminer les pullulations des simples variations saisonnières auxquelles sont soumises les populations de la plupart des espèces.

Cette dynamique explosive est associée à des déplacements massifs d'individus, qualifiés « d'invasion », depuis l'épicentre de pullulation vers de nouvelles aires, quelque fois caractérisées par des milieux écologiquement différents de l'aire d'origine. La faculté à coloniser de nouveaux milieux est liée à une grande tolérance écologique (Mason 1987), notamment pour les ressources alimentaires. Par exemple, les chenilles du bombyx disparate, *L. dispar*, qui présentent une nette préférence pour les chênes, peuvent en cas de pullulation coloniser les peuplements de hêtres, charmes, peupliers, bouleaux, saules, érables, tilleuls, aulnes mais aussi pins, épicéas, mélèzes, et sapins de Douglas, ainsi que la végétation arbustive ou herbacée. De même, les nuages de criquets migrateurs dénudent des paysages entiers, avec des dégâts enregistrés principalement sur Graminées (mil, sorgho, riz, maïs, canne à sucre, blé, orge…) mais également sur de nombreuses autres plantes (bananiers, ananas, palmier à huile, cocotier, etc) (COPR 1982). La faculté à coloniser de nouveaux milieux est également liée à de grandes capacités de dispersion. Ainsi, les chenilles du premier stade de *L. dispar* portent à la fois des poils longs (parfois plus longs que le corps) et des poils courts (portant une petite vésicule d'air) qui favorisent la dispersion par le vent, parfois sur plusieurs dizaines de kilomètres. Acridiens macroptères, les locustes sont des bons voiliers qui, en plus de se disperser par les vents dominants, forment en période de pullulation des nuages d'ailés dont la cohésion est maintenue activement par un comportement grégaire.

L'explosion des densités des populations des insectes phytophages et l'expansion de leurs aires d'habitat ont des influences délétères sur le bien-être ou la survie humaine, affectant notamment les écosystèmes agricoles et forestiers gérés par l'homme. *D. frontalis* est l'agent de mortalité des pins le plus important au Sud des Etats-Unis, au Mexique et dans certaines parties d'Amérique centrale. En Alabama, par exemple, le coût lié à ses défoliations peut atteindre jusqu'à 26 M$ par an (http://www.bugwood.org/barkbeetles/spb/index.HTML). Le criquet pèlerin, *Schistocerca gregaria*, est un locuste redouté, dont les essaims peuvent atteindre le sud de l'Europe, l'Afrique au nord de l'équateur et la péninsule Arabique et Indo-Pakistanaise. En 2003-2005, cette espèce a une fois de plus envahi l'Afrique sahélienne entraînant un coût d'environ 300 M$ (M. Lecoq, CIRAD, comm. pers.). C'est d'abord

la volonté de limiter les dégâts économiques et humains occasionnés par les insectes pullulants qui a généré un intérêt scientifique sur ces organismes, au travers de programmes d'écologie et d'entomologie appliquées. Ces disciplines ont déplacé leurs efforts de la destruction curative chimique, économiquement et environnementalement non supportables, au contrôle et/ou à la prévention des pullulations des insectes nuisibles. Ces nouvelles stratégies de lutte, reposant sur l'anticipation des pullulations et/ou sur une forte spécificité vis-à-vis de l'organisme d'intérêt, nécessitent, pour être efficaces, une compréhension approfondie de la biologie des populations des espèces d'intérêt et plus particulièrement de leur dynamique temporelle et spatiale (Thomas 1999 ; Lomer *et al.* 2001 ; Hunter 2004). Ces approches soulèvent également une question fondamentale, qui rejoint les questionnements de l'écologie théorique : pourquoi certaines populations pullulent dans certains lieux plus ou moins régulièrement, alors que dans d'autres populations, d'autres espèces et d'autres lieux, ces éruptions sont irrégulières ou n'arrivent jamais?

### 1.2. Déterminants extrinsèques aux populations

La recherche des facteurs à l'origine des fluctuations de densités des populations a débuté avec les travaux empiriques de Collett (1912) et Elton (1924) sur les petits mammifères et avec la formalisation théorique de Lotka (1925) et Volterra (1926) (dans Stenseth et Ims 1993). Ces publications constituent le point de départ de la dynamique des populations croisant l'observation et la collection sur du long-terme de données écologiques avec des modèles mathématiques et statistiques. Ce corpus de recherche a mis en avant les déterminismes écologiques extrinsèques aux populations comme facteurs causaux des dynamiques populationnelles fluctuantes. Les premières causes qui ont été évoquées pour expliquer les variations périodiques de tailles de population observées chez de nombreuses espèces sont les variations climatiques (Elton 1924 sur le lièvre, le lemming, et le lynx par exemple). Chez les insectes forestiers et prairials, il a été montré que de nombreuses espèces pullulent à la suite d'événements climatiques 'stochastiques', en particulier de sécheresse (revue dans Mattson et Haack 1987) ou de fortes pluies (Roffey et Popov 1968). Il semblerait que la sécheresse peut provoquer des pullulations d'insectes phytophages parce qu'elle procure un environnement physique plus favorable, mais également parce qu'elle modifie la qualité nutritive ou la toxicité des plantes (revue dans Mattson et Haack 1987). Dans ce dernier cas, c'est la relation

plante-herbivore, *via* la sécheresse, qui régule directement la dynamique des populations. De telles relations trophiques spécifiques, comprenant également les relations proie-prédateur (*e.g.*, *D. frontalis* Turchin 2003) et hôte-parasite (*e.g.*, *L. dispar* Elkinton *et al.* 1995), ont été privilégiées comme mécanismes déterminant les pullulations, dès l'émergence de la théorie de la régulation des populations densité-dépendante (Nicholson 1933, 1954).

Ces voies de recherches ont orienté le contrôle des organismes nuisibles vers l'utilisation d'agents biologiques influant sur la dynamique des populations nuisibles. Par exemple, l'importance du cortège d'ennemis naturels et notamment d'une virose, la polyhédrose nucléaire, dans la multiplication et la survie des populations de papillons forestiers (*e.g.*, la tordeuse de l'épinette *Choristoneura fumiferana* et le bombyx disparate *L. dispar*) a amené les gestionnaires à développer des stratégies basées sur de tels agents biologiques.

### 1.3. Déterminants intrinsèques aux populations

L'observation de différences phénotypiques ou génétiques entre les individus des périodes de pullulation et de rémission a conduit de nombreux auteurs à se tourner vers la recherche de mécanismes intrinsèques densité-dépendants. Le champ d'étude est ici déplacé de la communauté et ses interactions à la population et son polymorphisme. Les changements densité-dépendants de la population sont pertinents pour l'étude des pullulations uniquement s'ils concernent des traits affectant la dynamique des populations directement (*e.g.*, la dispersion, la fécondité et la survie) ou indirectement (*e.g.*, la compétition intra-spécifique, la structure en âge ou stades, le comportement de groupement, de coopération ou de territorialité). Ces traits sont classiquement regroupés sous le terme de « phase». Cette vue de la régulation des populations par une dépendance phasaire, abondamment développée chez les petits mammifères (revue dans Stenseth *et al.* 1998), n'a reçu qu'une attention récente chez les insectes (*e.g.*, Blackith et Albrecht 1979 ; Baltensweiler 1984 ; Rossiter 1992 ; Ginzburg et Taneyhill 1994 ; Rossiter 1994 ; Rossiter 1996 ; Ginzburg 1998). Les cas les plus frappants sont les effets d'agrégation, comme l'évoque la terminologie « attaques en masse de scolytes » et « nuages de criquets ». Lorsque les pins produisent de la résine, *D. frontalis* est incapable d'utiliser leur tissus pour se nourrir et se reproduire. Lors d'une attaque d'un

arbre, ce scarabée émet une phéromone d'agrégation qui attire les autres scarabées conspécifiques, jusqu'à atteindre des effectifs (environ deux milliers) tels que l'arbre est littéralement drainé de ses ressources en résine, rendant nulle sa capacité à se défendre (Wood 1972 ; Hodges *et al.* 1979). De même les populations 'grégaires' de locustes concentrent leurs effectifs lors des périodes de pullulation par un comportement de forte mobilité et de cohésion active avec les congénères (Roffey et Popov 1968 ; Despland *et al.* 2000).

Cette nouvelle voie de recherche a permis de diriger le contrôle des organismes nuisibles vers l'altération des traits de population impliqués dans la capacité à pulluler. Ces traits peuvent être manipulés (i) en remplaçant la population ravageuse par des individus génétiquement modifiés pour ces traits (Hoy 1994), ou (ii) en inhibant l'expression des traits impliqués dans la pullulation le plus classiquement par l'utilisation de substances chimiques impliquées dans le transfert d'informations lors d'une interaction entre deux individus. Récemment, l'exploitation de phéromones anti-agrégation a permis d'inhiber le comportement d'agrégation des scarabées des pins *D. frontalis* et a démontré un fort potentiel pour le contrôle de cette espèce au moins à un stade précoce de l'infestation (Borden 1989 ; Billings *et al.* 1995). Le phénylacétonitrile, impliqué dans la communication phéromonale chez les locustes, est actuellement à l'essai chez *S. gregaria*. Utilisé en pulvérisation sur des bandes larvaires de cette espèce, cette phéromone modifie le comportement des bandes et entraîne leur désagrégation à court terme (Hassanali et Bashir 1999).

### 1.4. Apports potentiels de la génétique des populations

Le rôle de traits populationnels dans le contrôle des pullulations d'insectes rend importante l'étude de la variation phénotypique observée pour ces traits au sein et entre les populations. Les sources potentielles de cette variation phénotypique proviennent du génotype dérivé des gènes transmis par les parents et des conditions environnementales contemporaines à la vie de l'individu et/ou des parents. Ainsi, le changement phasaire de la population avec les épisodes de pullulation peut être une réponse plastique aux changements environnementaux associés aux pullulations (*e.g.*, environnement abiotique, ennemis naturels). Un changement phasaire peut également provenir d'effets parentaux génétiques (*i.e.*, transmission des gènes nucléaires et/ou mitochondriaux par les parents; Chitty 1967) ou environnementaux (*e.g.*, transmission d'hormones ou de

micro-organismes symbiotiques; Rossiter 1996; Ginzburg 1998). Dans ces cas-là, la variation des traits phasaires est héritée, c'est-à-dire transmise d'une génération à l'autre. Sous cette condition, la réponse phasaire densité-dépendente peut être un facteur causal des pullulations. Etant donné les différences drastiques de sélection nécessaires pour produire un fort changement de fréquences génétiques en quelques générations, l'hypothèse d'une cause génétique des fluctuations a jusqu'à présent reçu peu de crédit (Mitter and Schneider 1987 ; cependant voir Baltensweiler 1984 sur la tordeuse du mélèze). En revanche, il est de plus en plus évident que les effets des environnements parentaux sont capables d'altérer la qualité de la population (Rossiter 1992) et de largement affecter la dynamique des populations (Ginzburg et Taneyhill 1994 ; Rossiter 1994 ; Beckerman 2002).

La génétique des populations vise à déterminer l'ampleur de la variation génétique au sein et entre les populations naturelles. En confrontant les niveaux et la distribution de la variation génétique à l'histoire des habitats (*e.g.*, les glaciations), à l'histoire de vie (*e.g.*, la capacité à diapauser) et/ou à l'environnement (*e.g.*, la température mensuelle la plus basse), la génétique des populations permet d'expliquer l'origine et le maintien de la variation génétique des populations. De même, la comparaison des patrons de variation de traits phénotypiques (*e.g.*, la propension à pulluler ou à grégariser), de gènes et de variables environnementales permet d'évaluer les parts génétique et environnementale de la variation phénotypique. Enfin, la génétique des populations permet d'analyser l'origine évolutive neutre et/ou sélectionnée de la variation génétique observée. Des modèles théoriques ont été développés pour estimer l'influence des forces évolutives majeures, neutres - à savoir la dérive génétique, la migration, et la mutation - et sélectives. La théorie neutre de Kimura soutient ainsi que le niveau de diversité génétique est déterminé par l'équilibre entre la mutation neutre, qui crée de la variabilité, et la dérive génétique, qui cause la fixation des allèles (David et Samadi 2000). La théorie de génétique des populations inclut également les effets, en terme de différenciation génétique, de la sous-division d'une population en groupes qui ne se croisent pas librement et des flux de gènes. Dans les théories sélectives, la sélection naturelle, par l'augmentation en fréquence des mutations avantageuses et la diminution des variants défavorables, crée un changement évolutif (David et Samadi 2000).

### 1.5. Mesurer la variation génétique neutre et sous sélection

Bien que des espèces montrent des polymorphismes génétiques visibles, pour lesquels les différences morphologiques sont déterminées par des génotypes à un ou quelques locus (*e.g.*, la couleur de l'œil humain, la forme des coquilles d'escargots), la variation est généralement 'polygénique' et les effets des locus individuels sont peu distinguables. Dans ce contexte, la biologie moléculaire est une discipline permettant d'accéder au polymorphisme génétique à de multiples locus, en particulier *via* l'utilisation de la réaction de PCR (Polymerase Chain Reaction), cette dernière permettant d'accéder à la variation de séquence ADN. L'utilisation de marqueurs héritables, discrets et stables permet d'identifier les génotypes qui caractérisent les individus et les populations.

L'étude de la variation de séquences non-codantes est informative quant aux niveaux et à la distribution de la variation génétique neutre des populations naturelles. Un niveau élevé de polymorphisme est nécessaire pour estimer la variation génétique au sein et entre les populations d'insectes pullulants puisque ces organismes ont de grandes capacités de dispersion qui génèrent probablement des flux de gènes importants entre les populations à de larges échelles géographiques. De bons marqueurs candidats sont alors les locus microsatellites qui ont de nombreux allèles, différant par un nombre variable d'éléments de séquences de 2 à 6 bases répétés en tandem, et de fortes hétérozygoties (revue dans Estoup et Angers 1998). A notre connaissance, l'étude de la variation de séquences microsatellites n'a été cependant que rarement appliquée aux insectes pullulants (voir cependant Bogdanowicz *et al.* 1997 sur *L. dispar*).

Les biologistes montrent un intérêt croissant pour la variation de gènes codant pour des traits d'intérêt (*i.e.*, potentiellement sous sélection). Accéder à cette variation est difficile, malgré les récents développements en génomique des populations (Luikart *et al.* 2003). Cette difficulté vient du fait que les traits d'intérêt sont déterminés par des facteurs génétiques multiples. Par exemple, les bases moléculaires du changement phasaire des locustes sont complexes, avec plus de 500 différences d'expression génique identifiées entre des locustes élevés isolés ou en groupes (Kang *et al.* 2004). Une alternative à l'isolement de(s) gène(s) d'intérêt, communément utilisée pour étudier la variation génétique de caractères sous sélection, consiste à étudier directement l'expression du trait d'intérêt en conditions environnementales contrôlées. Les traits de

phase ont d'ailleurs largement été étudiés en expérimentation contrôlée, tout du moins sous un angle physiologique, en particulier chez les locustes (revue dans Pener 1991).

## 2. LE CRIQUET MIGRATEUR, *LOCUSTA MIGRATORIA*

### 2.1. Systématique et Biogéographie

Parmi les milliers d'espèces d'Acrididae, cinq cent espèces environ peuvent pulluler et causer des dégâts à l'agriculture dans des conditions d'environnement favorables, mais seul un petit nombre d'entre elles, parmi les plus dangereuses, sont capables de présenter une transformation de phase complète au cours des périodes de pullulation : ce sont les locustes. La propension à changer de phase et à pulluler est cependant variable au sein des locustes : les deux espèces les plus ravageuses, le criquet pèlerin *S. gregaria* et le criquet migrateur *L. migratoria*, manifestent une importante grégarisation à partir d'une densité relativement faible (500 et 2000 ailés par hectare, respectivement ; Franc et al. 2005) alors que pour le criquet marocain, *Dociostaurus maroccanus*, dont l'impact économique est plus limité, seules des différences morphologiques ont été mises en évidence au cours des périodes de forte pullulation (Uvarov 1966). Seule la capacité à changer de phase singularise les locustes au sein de leur famille. A noter qu'ils sont présents dans différentes sous-familles d'Acrididae, et ne forment donc pas un groupe taxonomique distinct (Fig. 1).

Alors que toutes les autres espèces de locustes sont limitées aux régions les plus chaudes du globe, *L. migratoria* est répandu dans tout l'Ancien Monde. De nombreuses sous-espèces ont été décrites, sur la base de mesures de ratios morphométriques d'individus grégarires et/ou solitaires, généralement dans des aires géographiques où l'espèce pullule historiquement (Fig. 2). Le statut de ces sous-espèces et leurs limites géographiques précises demeurent cependant incertains. Ainsi, et à titre d'exemple, Zolotaresky (1929) a défini la sous-espèce *L. m. capito* regroupant les locustes malgaches, statut cependant mis en doute par Wintrebert (1970). Meyen (1835) et Uvarov (1936) ont reconnu la sous-espèce *L. m. manilensis* pour les locustes pullulant en Philippines, à Bornéo, en Malaisie et en Chine, nomenclature récemment révisée par Chen (1991) qui a étendu cette sous-espèce à l'Indonésie. La sous-espèce du pourtour méditerranéen (*L .m. cinerascens*) chevauche en de nombreux endroits l'aire de *L. m.*

**Fig. 1. Répartition taxonomique des 10 espèces de locustes dans les principales sous-familles d'Acrididae.** Ces relations taxonomiques sont tirées de Launois (1978).

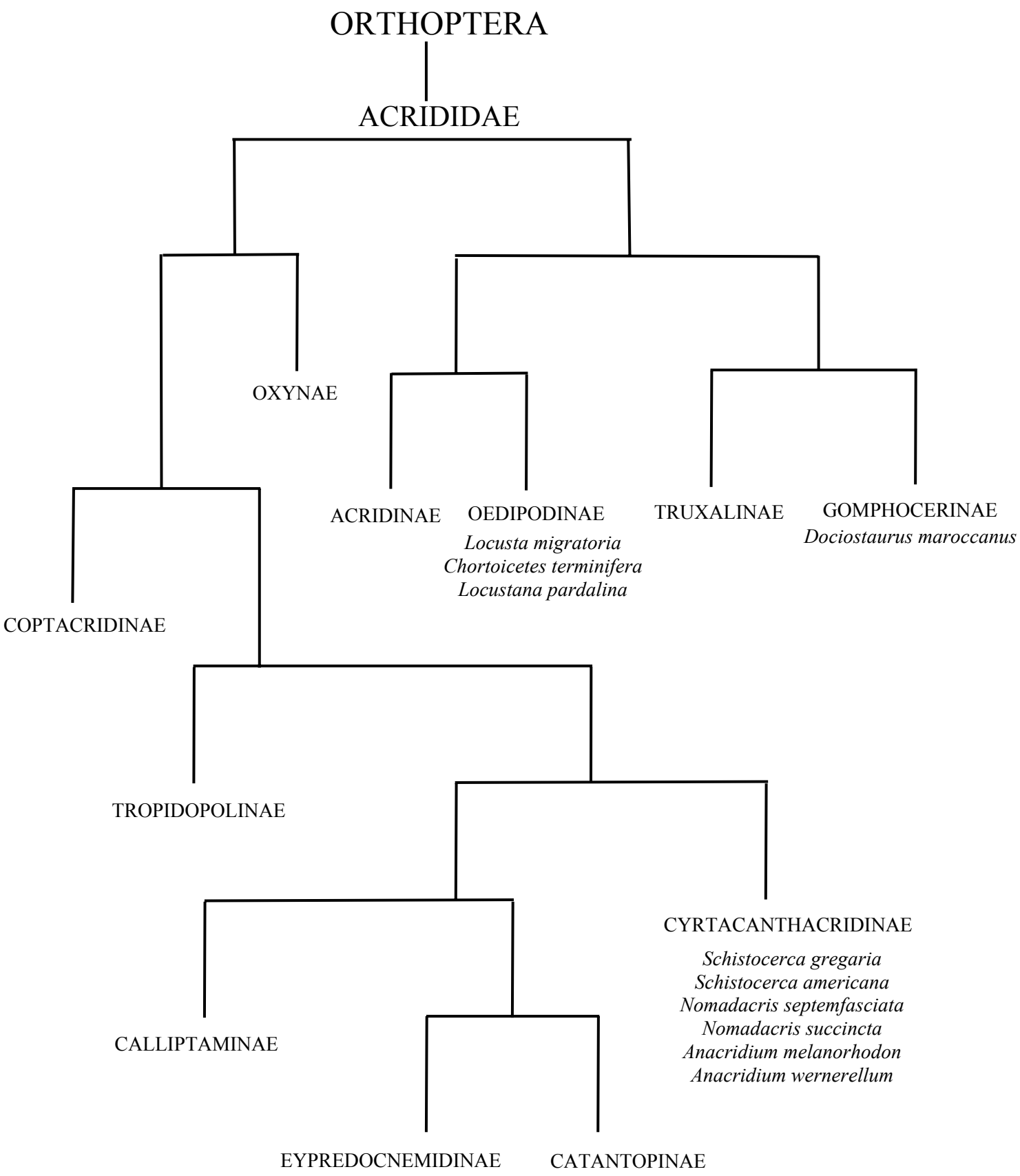
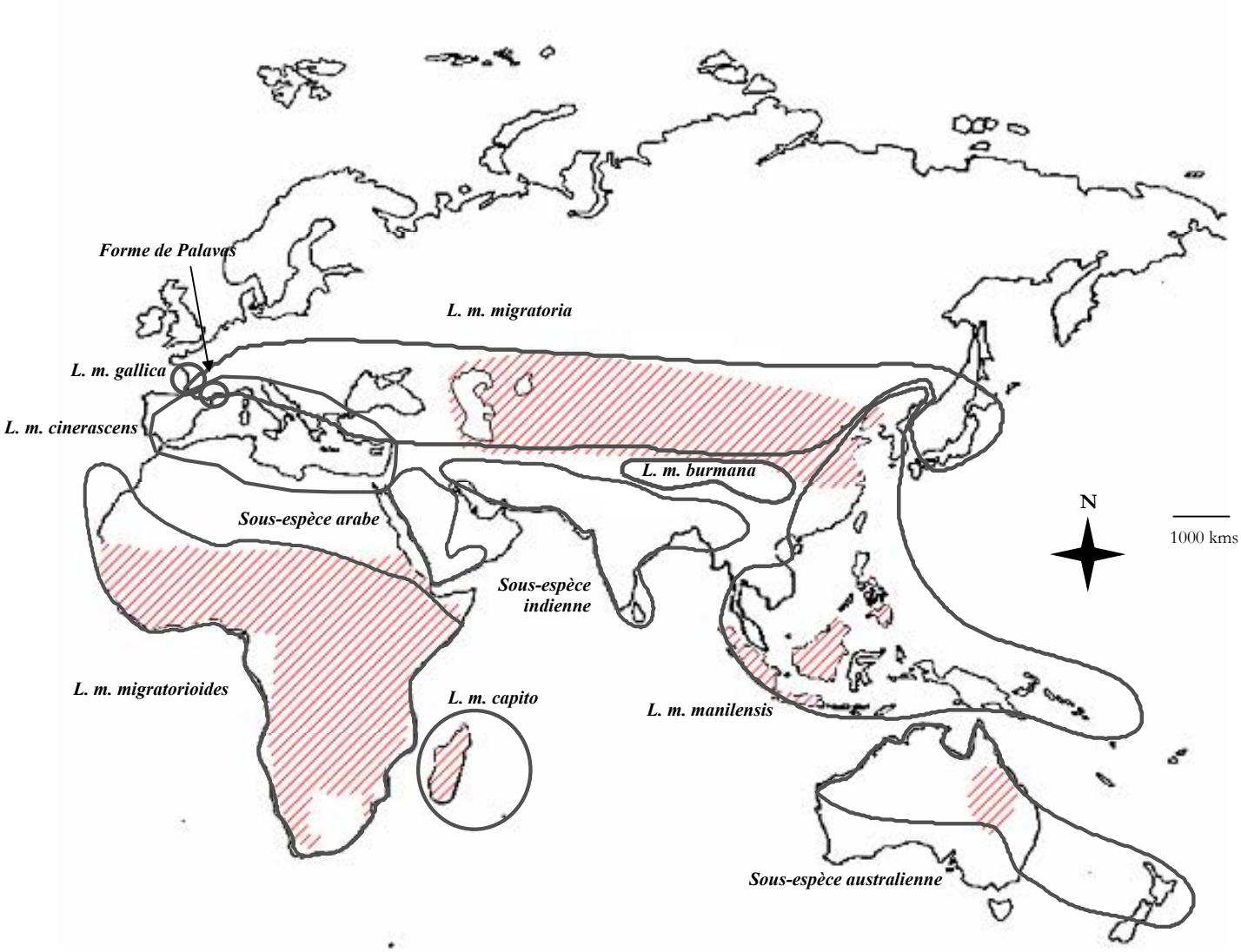


**Fig. 2. Répartition géographique des onze sous-espèces de *L. migratoria* (traits gris) et représentation sommaire des aires occupées par des populations fortement pullulantes (hachurées en rouge).**



*migratoria*. Une forme dite « de Palavas » étroitement localisée dans la région de l'Hérault, en France, a également été identifiée (Remaudière 1940a) et semble isolée au sein de l'aire de distribution de *L .m. cinerascens*. Les sous-espèces tibétaine, *L. m. burmana*, et française, *L. m. gallica*, ont également été identifiées (COPR 1982 ; Remaudière 1947). Des sous-espèces indienne, australienne et arabe ont été reconnues mais jamais nommées (COPR 1982). Farrow et Colless (1980) ont récemment proposé, sur des critères morphologiques comparables, de placer l'ensemble des populations tropicales dans une même sous-espèce, *L. m. migratorioides*, et les locustes du Paléarctique dans la sous-espèce *L. m. migratoria*. Cette proposition n'est cependant pas passée dans l'usage et la communauté scientifique continue actuellement à reconnaître les 11 sous-espèces mentionnées plus haut.

*L. migratoria* s'est adapté à des régions écologiquement très contrastées. Cette espèce, essentiellement graminivore, fréquente les formations herbeuses, de préférence hautes et denses de type savanes, cultures céréalières ou dunes littorales, mais également basses et ouvertes, de type steppes et prairies voire oasis. Les facteurs abiotiques sont les facteurs clés de l'écologie de *L. migratoria*. En milieu tropical, le facteur hydrique détermine la qualité de la végétation, et l'humidité du sol, nécessaires au bon développement et au succès de reproduction de cette espèce hygro-mésophile. *L. migratoria* s'est adapté à l'instabilité saisonnière des conditions écologiques, et notamment des pluies, par des déplacements sur de grandes distances (plusieurs centaines de kilomètres), tant en saison des pluies qu'en saison sèche, à la recherche de meilleures conditions écologiques (Farrow 1990 ; Lecoq 1975). En milieu tempéré, le développement de *L. migratoria* est également contrôlé par la température, avec une diapause embryonnaire facultative en saison froide. En conséquence, les individus de milieux tropicaux ont un développement continu, avec 3 à 5 générations par an (*e.g.*, Lecoq 1975), tandis qu'en milieu tempéré, le développement peut-être interrompu, avec 1 à 3 générations par an (*e.g.*, Roubaud 1947 ; Ma 1958). L'impact des parasites et des prédateurs semble négligeable, au moins en milieu tropical du fait de la variabilité saisonnière des conditions environnementales et des migrations des locustes (Roffey et Popov 1968; Wintrebert 1970; Farrow 1974).

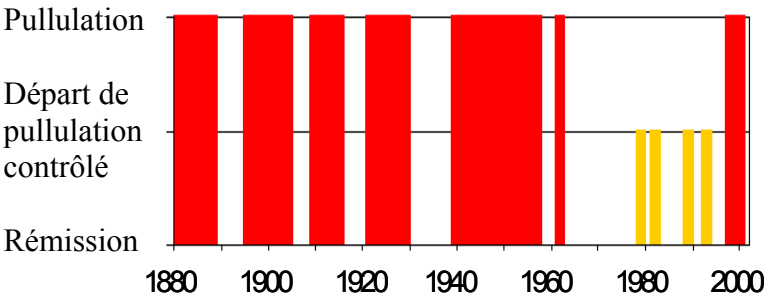
### 2.2. Variation géographique de la propension à pulluler

La propension à pulluler de *L. migratoria* varie géographiquement (Fig. 2). *L. migratoria* a toujours sévit à Madagascar et en Asie. Ainsi, 7 et 16 évènements majeurs de pullulation ont été respectivement enregistrés dans ces deux zones (Randriatmanantsoa 1998 ; Zhang et Li 1999) (voir la Fig. 3 qui illustre l'importance des pullulations à Madagascar). L'Europe centrale est également touchée, même si les pullulations y sont moins fréquentes et importantes (Waloff 1940). L'Afrique sub-saharienne a été historiquement très fortement touchée par les essaims de *L. migratoria*, mais depuis l'endiguement récent du fleuve Niger et la conséquente modification de la dynamique des crues dans l'aire grégarigène de l'espèce au Mali, plus aucune pullulation n'a été reportée. L'Afrique du Nord et l'Indonésie sont nouvellement concernées par les pullulations de *L. migratoria*, depuis que le changement de l'utilisation des terres a ouvert des milieux favorables à l'espèce. A Sumatra, par exemple, les populations sources de pullulations récentes et généralisées semblent provenir des nouvelles plantations de canne à sucre mises en place à la suite de la déforestation intensive des dernières décades (Lecoq et Surkino 1999). Dans le Sahara algérien, des pullulations récentes, d'importance toutefois modérée, de *L. migratoria* ont été observées, corrélées à l'expansion récente de cultures céréalières irriguées (*e.g.*, blé dur, *Triticum durum* et sorgho *Sorghum vulgare* ; Benfekih *et al.* 2002). L'Europe de l'ouest n'est plus affectée de manière significative par *L. migratoria* depuis plusieurs siècles : le dernier événement de pullulation, dont l'origine putative est la Mer Noire, remonte au XIV$^{ème}$ siècle (Waloff 1940). Depuis, seules quelques pullulations locales ont été recensées (*e.g.*, région de Naples en 1936; Jannone 1947; région des Landes, en 1946-1948; Glize 1996).

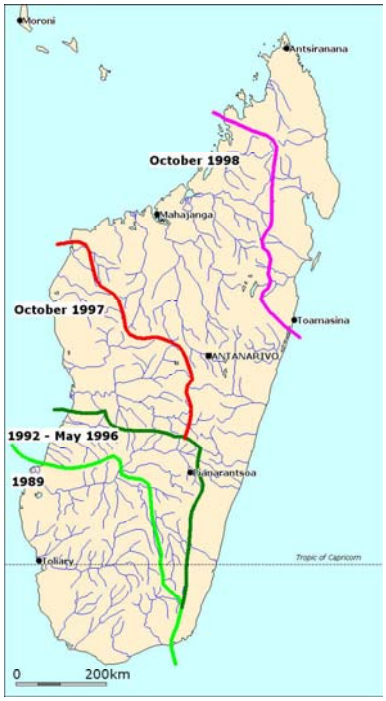
Les déterminants environnementaux des pullulations sont relativement bien compris à l'échelle de la région écologique des populations concernés. En milieu semi-désertique africain et malgache, les augmentations numériques des populations au début des pullulations sont associées à de fortes pluies, procurant une végétation abondante et un sol humide permettant à des vagues successives de femelles de pondre et diminuant la dessiccation des œufs et des larves (Davey 1956; Waloff 1962; Dempster 1963; Têtefort et Wintrebert 1966; Roffey et Popov 1968 ; Launois 1974). Dans le delta central du fleuve Niger au Mali, les niveaux d'humidité favorables aux départs de pullulation sont associés au phénomène de décrue des fleuves et des lacs (Farrow 1975).

**Fig. 3. Pullulations de *L. migratoria* à Madagascar : historique des pullulations (a), aire d'invasion de la dernière pullulation (b) et photographies d'essaims d'adultes et de bandes larvaires (c).**

**(a)**

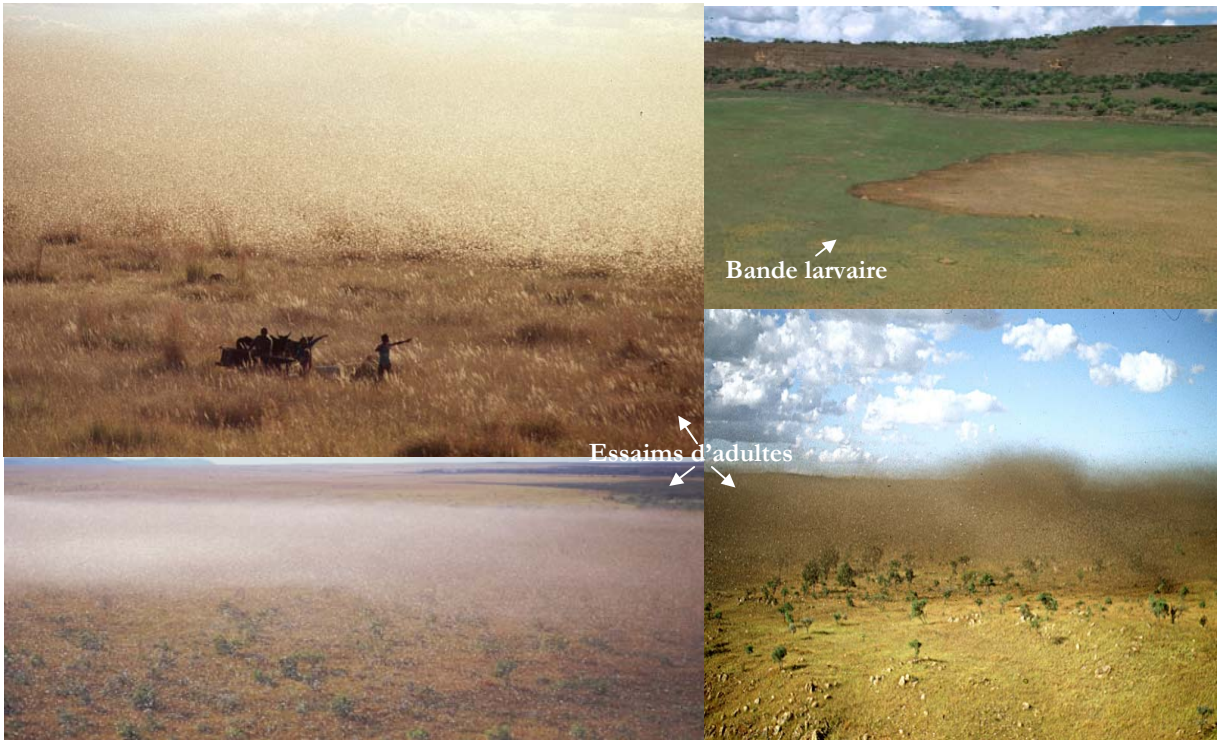



**(b)**




Duranton et al. 2000

**(c)**




Photos : Mr Duranton (bande larvaire) et Mr Lecoq (essaims), CIRAD, UPR Acridologie, Montpellier

En revanche, en milieu tropical indonésien et philippin, excessivement humide pour *L. migratoria*, les évènements de pullulation sont corrélés avec des périodes de sécheresse (COPR 1982 ; Lecoq et Surkino 1999). Enfin, en milieu tempéré chinois, les pullulations sont corrélés avec les épisodes de réchauffement et sécheresse associés à El Niño (Zhang et Li 1999). A une échelle locale, il a été montré que la distribution agrégée des ressources (Despland *et al.* 2000) et une forte variation du microclimat (Farrow 1975) concentrait les individus, et de ce fait facilitait les départs de pullulations. Il est important de noter que les différences de propension à pulluler corroborent également avec la classification taxonomique actuelle réalisée sur la base de critères morphométriques (Farrow et Colless 1980 ; Fig. 2). Ainsi *L. m. migratorioides*, *L. m. capito*, et *L. m. manilensis* sont des sous-espèces pullulantes, alors que *L. m. gallica*, *L. m. cinerascens*, *L. m. burmana*, les sous-espèces arabes et indiennes, et la forme de « Palavas » sont non pullulantes. Le statut de pullulations des sous-espèces australienne et *L. m. migratoria* est plus obscur puisque certaines de leurs populations semblent pulluler et d'autres non.

### 2.3. Les phases solitaire et grégaire des périodes de rémission et pullulation

Cette partie expose les clefs du changement phasaire chez les locustes nécessaires à la compréhension du travail de thèse. Elle est cependant non exhaustive ; pour plus de détail, il faut se référer aux revues de Pener (1991), Pener et Yerushalmi (1998), Breuer *et al.* (2003), et Simpson *et al.* (2006).

#### *2.3.1. Différences entre les deux phases et leur quantification*

Les changements de phase chez les locustes sont particulièrement impressionnants, comme l'illustre la croyance jusqu'en 1921 que les phases solitaire et grégaire de *L. migratoria* étaient deux espèces différentes, *L. danica* et *L. migratoria* (Uvarov 1921). Les changements phasaires induisent en effet chez *L. migratoria* des différences profondes sur un grand nombre de caractéristiques qui incluent la morphologie, le comportement, l'écologie, l'immunité et la physiologie. Les traits morphologiques et comportementaux ont fait l'objet de la majorité des recherches et sont aujourd'hui précisément quantifiés chez *L. migratoria* (*e.g.*, Dirsh 1953; Hoste *et al.* 2002). Le

changement phasaire le plus frappant est un changement chromatique : chez *L. migratoria*, les individus solitaires sont monochromes, vert ou marron clair selon le substrat, alors que les individus grégaires sont caractérisés par un dépôt en mélanine sur une robe jaune (Nolte, 1973) (Fig. 4). Les traits les plus classiquement utilisés en acridologie pour mesurer l'état phasaire, grégaire ou solitaire, sont des rapports morphométriques, au nombre d'une vingtaine (Lauga 1977). Il est par exemple bien connu que les individus solitaires ont un pronotum convexe alors que les individus grégaires ont un pronotum plat, cette différence morphologique étant détectable à l'œil (Fig. 4) et mesurable par le rapport de la hauteur maximale sur la longueur du pronotum. Sur le terrain, les différences de comportement entre les deux phases sont évidentes, puisque les individus grégaires s'attirent, s'agrègent, sont très mobiles et actifs, et pour les imagos, volent de jour, alors que les individus solitaires sont cryptiques, dispersés, relativement inactifs, et volent essentiellement en début de nuit (Fig. 4). La description approfondie du comportement phase-spécifique des locustes et la mise en place d'une méthodologie de mesure de la phase comportementale sont cependant relativement récentes (revue dans Simpson *et al.* 1999).

### *2.3.2. Variations environnementales et génétiques*

Il est bien connu depuis la découverte du changement de phase chez les locustes que les conditions de groupement expérimentées par un individu ($E_O$) induisent la grégarisation (Fig. 5 ; Uvarov *et al.* 1936). Cependant, Islam *et al.* (1994a) et Bouaïchi *et al.* (1995) ont les premiers montré que la densité expérimentée par les parents ($E_M$) influence également le degré de grégarisation du descendant chez *S. gregaria*. Par exemple, une vie pré-reproductive en isolement suivie d'une période d'oviposition en conditions de groupement conduit à une descendance au comportement et à la morphologie grégaires (Fig. 5). Lauga et Hatté (1977) et Albretch *et al.* (1959) ont montré des effets parentaux dépendant du groupement dans les valeurs de traits reproductifs du criquet migrateur *L. migratoria*. La présence d'une source environnementale parentale dans le changement de la phase comportementale solitaire à la phase grégaire a directement été démontrée par Mc Caffery *et al.* (1998) qui ont mis en évidence que des mères *S. gregaria* groupées au moment de la ponte ajoutaient à leur ponte un facteur grégarisant hydrophile. Finalement, Islam *et al.* (1994b) ont montré que l'impact de l'effet parental sur la phase comportementale de *S. gregaria* était

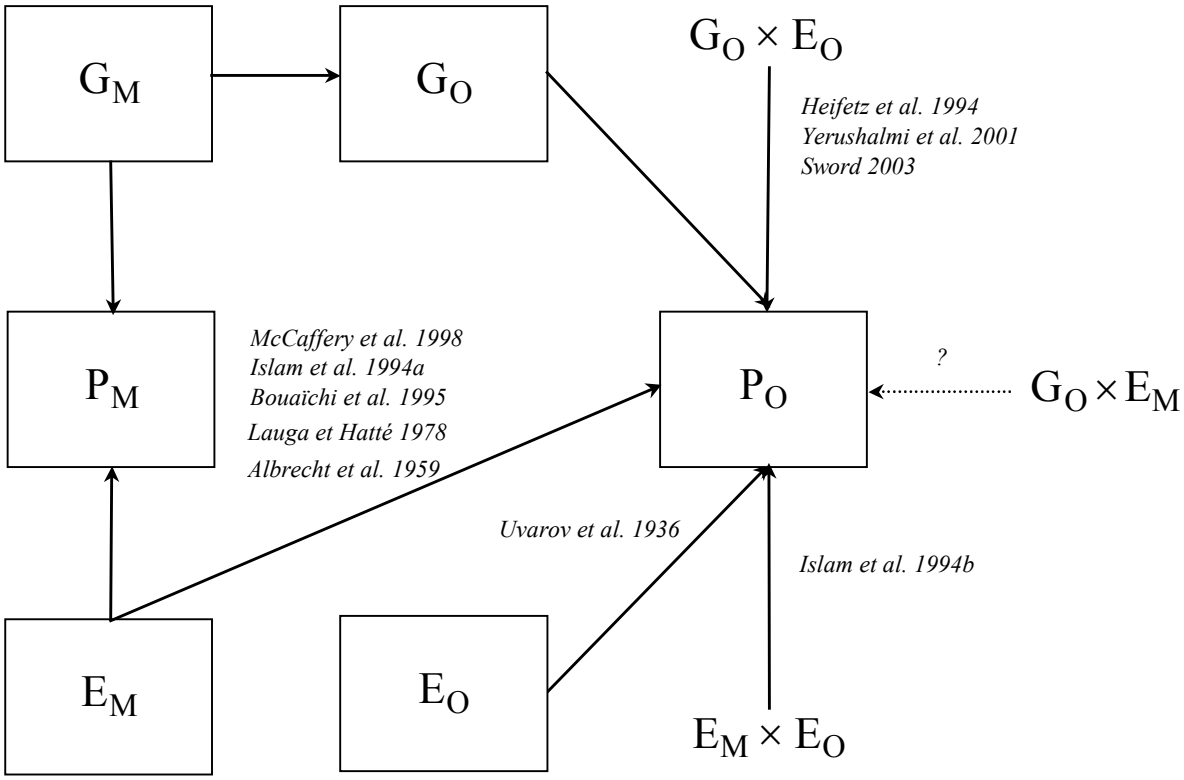
**Fig. 4. Photographies d'individus solitaires (a) et grégaires (b) de *L. migratoria* illustrant les différences pigmentaire, morphométrique et comportementale entre les deux phases.**

**(a)**


**(b)**


Photos : Mr Duranton et Mr Lecoq, CIRAD, UPR Acridologie, Montpellier

**Fig. 5. La phase d'un individu est déterminé par son propre génotype et environnement, mais également par son environnement maternel.** Ce modèle est tiré de celui de Rossiter (1998). Les sources potentielles de variation phénotypique du descendant, $P_O$, proviennent de sources génétiques et environnementales durant la génération parentale et éventuellement des générations antérieures à celle des parents (M) ou la génération du descendant (O). Ces sources incluent le génotype du descendant dérivé des gènes nucléaires transmis par les parents ($G_O$), les conditions environnementales contemporaines à la vie de l'individu ($E_O$), les conditions environnementales contemporaines à la vie des parents ($E_M$).



dépendant des conditions de densité expérimentées par la descendance, suggérant la présence d'une interaction $E_M$ x $E_O$ (Fig. 5).

Certains auteurs ont récemment cherché une variation génétique ($G_O$) du changement phasaire comportemental induit par des conditions d'isolement ou de groupement durant la vie de l'individu ($E_O$) dans le genre Schistocerca et chez *L. migratoria* (*i.e.*, l'interaction $G_O$ x $E_O$ ; Fig. 5). Bien que ces études aient montré des résultats peu clairs, elles tendent à supporter la présence d'effets $G_O$ x $E_O$ (Fig. 5 ; Heifetz *et al.* 1994; Yerushalmi *et al.* 2001; Sword 2003). Il est important de noter que la question de la variation génétique de la grégarisation héritée à travers les générations, indépendamment de la plasticité phénotypique exprimée au cours de la vie d'un individu, n'a encore jamais été adressée chez les locustes (*i.e.*, interaction $G_O \times E_M$; Fig. 5).

### *2.3.3. Facteurs proximaux*

Chez *S. gregaria*, le stimulus majeur de la grégarisation comportementale est le contact tactile, particulièrement au niveau des fémurs arrières (Hägele et Simpson 2000 ; Simpson *et al.* 2001). La combinaison de stimulus olfactif et visuel est également grégarisante au niveau comportemental chez cette espèce, mais ces stimulus ne sont pas efficaces seuls (Roessingh *et al.* 1998). Les stimulus impliqués dans les changements phasaires autre que le comportement ne sont pas bien connus.

Depuis les travaux de Joly (1951), un grand nombre d'études ont été entreprises sur *S. gregaria* et *L. migratoria* pour déterminer les facteurs physiologiques et moléculaires des différences phase-spécifiques morphologiques et comportementales des locustes (Pener 1991). Bien que ces études aient précisé le rôle d'un certain nombre de molécules (*i.e.*, la corazonine, les neuroparsines, certains inhibiteurs protéiques, etc), elles n'ont pu résoudre s'il y a ou non un inducteur moléculaire majeur du changement phasaire (Hassanali et al. 2005 ; De Loof *et al.* 2006). En particulier, la littérature semio-chimique est abondante mais confuse. Les molécules impliquées dans le comportement d'agrégation sont encore mal identifiées, et inversement, le rôle des molécules clairement identifiées est souvent sujet à controverse (revue dans Pener 1991 ; Pener et Yerushalmi 1998).

### *2.3.4. Facteurs évolutifs*

La question de la signification adaptative du changement phasaire est récente et encore balbutiante. La théorie la plus développée soutient, qu'au moins chez *S. gregaria*, les traits phasaires comportementaux et morphologiques fonctionnent en concert comme une stratégie anti-prédatrice (Simpson *et al.* 2006). Notamment, les patrons de coloration noire et jaune des nymphes de *S. gregaria* seraient une signalisation manifeste, avertissant les prédateurs potentiels que les individus ont mangé des plantes toxiques (Sword *et al.* 2000). Cette théorie est cependant moins plausible chez *L. migratoria* puisque cette espèce se nourrit essentiellement de graminées, généralement considérées comme peu riches en composés secondaires toxiques. Des études plus récentes suggèrent que l'agrégation, par des effets de confusion et de dilution du risque, protègent également les individus des prédateurs. Il a en effet été montré, chez le grillon *Anabrus simplex*, que les individus présents dans les bandes migratrices étaient moins prédatés que les individus hors de ces bandes (Sword *et al.* 2005). La grégarisation pourrait également conférer une meilleure résistance aux parasites et aux pathogènes par une sur-activation du système immunitaire chez *S. gregaria* (Wilson *et al.* 2002). Les différences de balance nutritive et de chimie phéromonale observées entre les deux phases de *S. gregaria* pourraient être une adaptation à des différences de compétition respectivement pour les ressources alimentaires et pour l'accès aux femelles entre les conditions de faible et de forte densités (Simpson *et al.* 2002 ; Seidelmann *et al.* 2005). Enfin, il semble que les changements de la morphométrie, du comportement et de la balance nutritive entre les phases solitaire et grégaire représentent une adaptation au comportement de migration (Uvarov 1966 ; Applebaum & Heifetz 1999 ; Ellis 1953).

### *2.3.5. Rôle dans la dynamique de populations*

Le changement de phase densité-dépendant est non seulement central à la biologie des locustes, mais également à leur importance comme espèces ravageuses. Seuls les locustes grégaires, rassemblés en bandes pour les larves et en essaims pour les adultes, forment des pullulations alors que les locustes solitaires vivent à faibles densités. La relation grégarisation – pullulation est aujourd'hui expliquée par le fait que les traits de migration et d'agrégation actives des individus grégaires ont pour conséquence de concentrer les individus, et par là d'affecter directement la dynamique des populations.

Par exemple, des facteurs environnementaux locaux et régionaux (voir section 2.2), telle que la distribution agrégée des ressources, favorise la congrégation passive des populations aux échelles locale et régionale. Sous ces conditions, les locustes alors en contact proche les uns des autres vont rapidement changer de phase et s'agréger activement (*e.g.*, Roffey et Popov 1968 ; Despland *et al.* 2000; Babah et Sword 2004). Ce mécanisme clé entraîne une rétro-action positive qui conduit à de plus grands groupements des individus, et donc à une plus forte agrégation.

### 2.4. Vers une nouvelle gestion des populations

Les pesticides chimiques appliqués par hélicoptère ont permis un contrôle relativement efficace des locustes en Afrique, Amérique du Sud, Chine et Australie depuis la fin de la seconde guerre mondiale (revue dans Hunter 2004). Mais au vu de la résistance économique et environnementale croissante à l'utilisation des pesticides chimiques, il est devenu urgent de développer des programmes de gestion intégrée, reposant sur une compréhension approfondie de la biologie des populations (Thomas 1999).

Une stratégie prometteuse, encore à ses débuts, est de traiter précocement *via* une approche du type lutte biologique. En effet, des produits alliant des agents de contrôle chimique et biologique (*e.g.*, la microsporidie *Nosema locustae* ou le champignon entomopathogène *Metarhizium flavoviride*) ont fait leur preuve au laboratoire et sur le terrain en Afrique, Australie, Chine, et Madagascar et promettent d'être particulièrement utiles pour traiter les environnements sensibles, tels que les habitats humides favorables à *L. migratoria* (revue dans Zhang et Hunter 2005). Comme l'action de l'agent biologique est retardée, elle nécessite d'agir précocement dans le processus de pullulation, *i.e.*, de contrôler les premiers rassemblements des locustes solitaires plutôt que les impressionnants essaims très mobiles de locustes grégaires (Lecoq 2001). Cette stratégie en cours de développement est actuellement limitée par la nécessité de déterminer l'origine des populations donnant naissance aux pullulations et leurs mouvements à venir. En effet, les approches directes, telles que les expérimentations de capture-marquage-recapture, sont limitées pour évaluer la structure et la dynamique des populations des locustes, du fait de la grande capacité de dispersion de ces organismes (Duranton *et al.* 1979). Ces spécificités incitent au développement d'autres méthodologies telle que la génétique des populations.

## 3. PRINCIPAUX OBJECTIFS DE LA THESE

L'objectif de cette thèse est d'apporter de l'information nouvelle sur le processus de pullulation chez les insectes phytophages par l'approche génétique peu explorée chez ces organismes, en considérant l'espèce *L. migratoria* comme modèle biologique.

Cette thèse s'articule autour de trois axes présentés dans trois chapitres :

(i) Caractéristiques et spécificités d'analyse des marqueurs microsatellites

(ii) Variation génétique populationnelle aux marqueurs microsatellites

(iii) Evolution expérimentale de la grégarisation

### 3.1. Caractéristiques et spécificités d'analyse des marqueurs microsatellites

Un premier objectif est d'évaluer le potentiel informatif des marqueurs génétiques en biologie des insectes pullulants. Cette motivation vient du fait que cette approche, et plus particulièrement l'emploi des séquences microsatellites, sont relativement neufs chez ces organismes. Une des raisons du manque d'études moléculaires sur les espèces pullulantes tient certainement à ce que ces insectes sont pour la plupart des lépidoptères et des orthoptères, pour lesquels la caractérisation et l'analyse des marqueurs microsatellites pose des difficultés. En effet, il a été montré que la faible fréquence de locus microsatellites, la forte prévalence d'allèles nuls à ces locus, et les duplications fréquentes de gènes rendent difficile le développement des marqueurs microsatellites chez les lépidoptères (revue dans Meglecz *et al.* 2004). Il est possible qu'au moins certains de ces mécanismes expliquent le faible nombre d'études de caractérisation d'amorces microsatellites chez les orthoptères (revue dans Zhang *et al.* 2003). Cinq des huit études de caractérisation d'amorces microsatellites chez les orthoptères rapportent des hétérozygoties observées inférieures aux hétérozygoties attendues à la majorité de leur locus microsatellites, suggérant de fortes fréquences d'allèles nuls dans cet ordre (King *et al.* 1998; Hockham *et al.* 1999; Bernardini *et al.* 2002; Zhang *et al.* 2003; Bailey *et al.* 2005a). Une utilisation pertinente des marqueurs microsatellites en génétique des populations des insectes pullulants requiert donc de mieux décrire les limites de caractérisation et d'analyse de ces marqueurs lorsque ceux-ci sont

caractérisés par des taux élevés d'allèles nuls, et si nécessaire de développer des méthodes appropriées.

### 3.2. Variation génétique populationnelle aux marqueurs microsatellites

Les approches de génétique des populations sont potentiellement pertinentes pour préciser la biologie des populations nuisibles, et en particulier pour informer sur le processus de pullulation.

Chez *L. migratoria*, certains auteurs remettent en cause la classification actuelle des onze sous-espèces avec des propensions à pulluler contrastées (*e.g.*, Farrow et Colless 1980). Il faut également noter que l'ensemble des taxonomistes ont basé leur travail sur des rapports morphométriques, originellement définis pour discriminer les phases des locustes. Cette technique inusuelle paraît injustifiée car la phase morphométrique des locustes est une réponse continue et en partie plastique aux densités de populations. Ainsi, la taxonomie de *L. migratoria*, et les facteurs qui la façonnent (*e.g.*, écologie, géographie, caractère pullulant), restent à ce jour largement incompris. La validité biologique de la classification actuelle de *L. migratoria* reste donc à être évaluée. Nous proposons dans cette thèse de réaliser une phylogéographie mondiale de *L. migratoria* à partir de marqueurs moléculaires, tels que les microsatellites. Cette discipline vise à révéler les processus historiques, géographiques et écologiques gouvernant la distribution géographique de la variation génétique (Avise 2000). Nous espérons ainsi reconnaître des groupes de populations avec des histoires évolutives plus ou moins indépendantes, et évaluer si ces groupes de populations sont congruents avec l'occurrence et l'absence géographique des évènements de pullulation.

Au vu de la demande des gestionnaires des populations pullulantes de locustes, nous chercherons également à préciser la structure et la dynamique des populations pullulantes de *L. migratoria* à une échelle plus locale à partir d'inférences basées sur les marqueurs microsatellites. L'analyse de marqueurs microsatellites permet en effet d'accéder à des processus évolutifs neutres et contemporains, telles que la migration et la dérive, et ainsi de renseigner sur des paramètres d'intérêt des populations, telles que les tailles de populations et les mouvements efficaces des individus entre populations. Les marqueurs microsatellites ont également montré un bon potentiel à déterminer les sources et les routes d'invasion (*e.g.*, Miller *et al.* 2005).

### 3.3. Evolution expérimentale de la grégarisation

*L. migratoria* est un modèle biologique de choix pour comprendre les déterminants des pullulations d'insectes. Entre autres, puisque certaines de ses populations sont clairement pullulantes alors que d'autres ne le sont pas et que sa répartition géographique englobe des régions aux environnements variés, cette espèce est un modèle idéal pour évaluer les parts environnementale et génétique de la capacité à pulluler des populations. Les études précédentes ont mis en évidence que la variation géographique de la propension à pulluler des populations de *L. migratoria* est liée à la variation géographique de facteurs climatiques extrinsèques. On peut cependant poser la question d'une variation géographique de la propension à pulluler également liée à une variation génétique de la capacité à pulluler des populations. Cette explication alternative, non exclusive de la précédente, est plausible car elle repose sur l'hypothèse que des traits intrinsèques aux populations participent à la formation ou au maintien des pullulations, ce qui a déjà été montré chez *L. migratoria* (Roffey et Popov 1968 ; Despland *et al.* 2000). Nous proposons donc ici d'évaluer la variation génétique de la propension à grégariser. Notons que cette question sera également adressée en confrontant la variation populationnelle de la propension à pulluler à la variation aux marqueurs microsatellites (voir section 3.2 de l'Introduction).

Par ailleurs, comme la phase de *L. migratoria* est aisément mesurable et qu'elle est fortement corrélée à la dynamique de populations, *L. migratoria* est un modèle prometteur pour étudier la phase-dépendance des pullulations d'insectes, et en particulier pour préciser le rôle de la grégarisation dans la dynamique de pullulation de cette espèce. Les variations phasaires comportementales, et entre autres d'agrégation, semblent jouer un rôle dans la dynamique de pullulation des locustes. Il n'est, en revanche, pas résolu si l'importance du processus de grégarisation dans la formation et/ou le maintien des pullulations est également liée à des traits non-comportementaux. Les traits de reproduction et de survie sont de bons candidats car leur variation affecte directement la dynamique des populations. Cependant, l'occurrence de changements majeurs de l'histoire de vie avec la phase a été peu étudiée chez les locustes, et est actuellement toujours sujette à controverse. Certains auteurs soutiennent que la grégarisation affecte profondément le potentiel de multiplication chez Locusta, Nomadacris et Schistocerca (Pener 1991), et d'autres mettent en évidence des effets maternels phase-dépendants plus complexes (Albrecht *et al.* 1959 ; Lauga et

Hatté 1978). Il faut noter que de la prudence s'impose vis-à-vis des résultats publiés sur les traits d'histoire de vie chez les locustes puisque les conditions expérimentales de mesures ainsi que le changement de phase n'ont généralement pas été contrôlés dans ces études. Nous rechercherons donc dans cette thèse si les traits d'histoire de vie de *L. migratoria* diffèrent avec la phase, par une expérimentation d'élevage permettant (i) de contrôler que les phases solitaire et grégaire ont été induites au cours de l'élevage et (ii) d'homogénéiser les conditions environnementales, et en particulier de densité, entre les deux phases comparées.

# **Chapitre I.** Caractéristiques et spécificités d'analyse des marqueurs microsatellites

L'objectif de ce chapitre est de préciser (i) les caractéristiques des marqueurs microsatellites, qui rendent leur développement et leur analyse difficiles chez *L. migratoria*, et en conséquence (ii) les spécificités d'analyse de ces marqueurs chez cette espèce. Ce dernier point concerne tout particulièrement la présence d'allèles nuls aux marqueurs microsatellites, très fréquemment reportée dans les études de génétique des populations ou de caractérisation des marqueurs microsatellites (Dakin et Avise 2004), et a donc une portée générale dépassant largement le cadre des insectes pullulants.

Ce chapitre débute par un avant-propos sur le choix d'utilisation des marqueurs microsatellites. Les résultats présentés par la suite concernent respectivement les patrons d'amplification multibandes et la présence d'allèles nuls observés aux marqueurs microsatellites. Pour plus de détail sur ces résultats, lire les manuscrits 1 et 2 ainsi que l'encadré 1. L'encadré 1 est une courte note complémentaire des informations présentées dans les manuscrits. Ce chapitre se réfère également à quelques résultats exposés dans les manuscrits 3 et 4.

## 1. CHOIX DES MARQUEURS MOLECULAIRES

La première étape de l'étude de la variation génétique des populations et des sous-espèces consiste à trouver des marqueurs présentant un niveau de variabilité adéquat pour le niveau de résolution souhaité. Les marqueurs microsatellites sont des marqueurs appropriés pour l'étude de la génétique des populations aux échelles géographiques locale et régionale (Estoup et Angers 1998). Bien que les marqueurs microsatellites permettent également de réaliser des études de phylogéographie intraspécifique (*e.g.*, Estoup *et al.* 1995), la majorité des études cherchant à résoudre les relations génétiques entre taxons potentiellement divergents chez les insectes utilisent des fragments d'ADN mitochondriaux (en particulier Cytochrome B, COI, et ADN ribosomique). Ces marqueurs ont un fort potentiel pour ce type d'études parce qu'ils ne sont pas sujet aux phénomènes de recombinaison et d'homoplasie, ne posent pas, comme les séquences nucléaires, de difficultés de lecture des séquences des génotypes hétérozygotes et ont une structure phylogéographique souvent plus marquée que les locus nucléaires.

Nous n'avons, malheureusement, pas réussi à développer avec succès ce type de marqueurs, ni d'autres marqueurs que les microsatellites. Plusieurs auteurs ont montré que les marqueurs mitochondriaux sont sujets à de l'assimilation nucléaire fréquente,

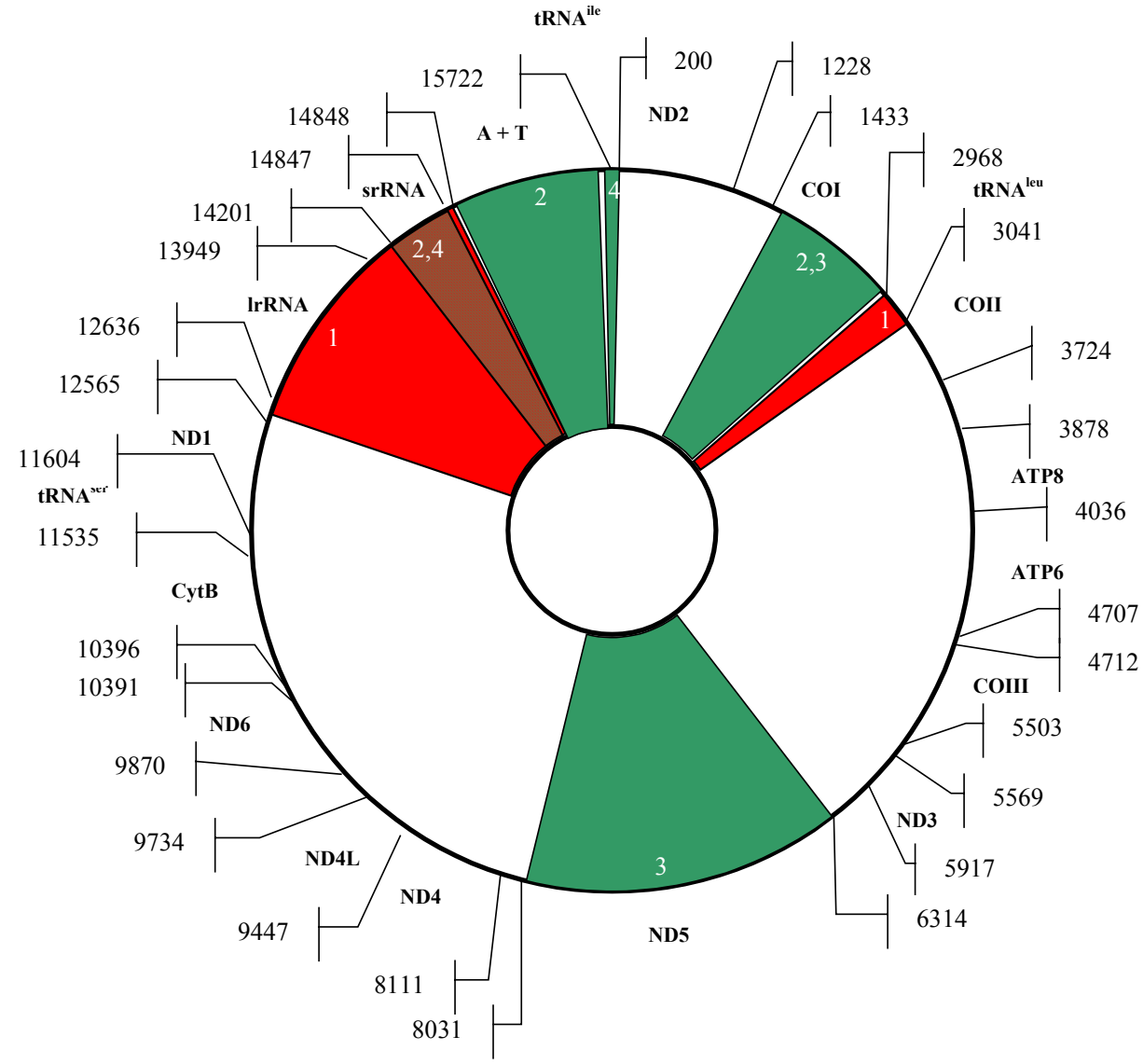
engendrant des copies nucléaires responsables d'ambiguités dans les inférences basées sur ces séquences. Les séquences nucléaires similaires à l'ADN mitochondrial sont particulièrement communes chez les criquets et ont été trouvées chez *L. migratoria* (Gellissen et Michaelis 1989 ; Zhang et Hewitt 1996 ; Bensasson *et al.* 2000; Fig. 6). D'autre part, l'analyse du polymorphisme de séquence d'un fragment du gène mitochondrial 12sRNA de trois sous-espèces de *L. migratoria* a révélé un faible niveau de variabilité (entrées GenBank ; Numéros d'accessions : AY324452 à AY324454). Ainsi, de par le risque de présence de copies nucléaires et d'absence de polymorphisme, nous avons renoncé à développer des marqueurs mitochondriaux et centré nos efforts sur des fragments d'ADN nucléaires, les espaceurs ribosomiques internes (ITS). L'amplification et la séquence des fragments d'ADN des gènes ITS1 et ITS2 (environ 800 paires de bases) ont été mises au point. Malheureusement, ces gènes se sont avérés monomorphes.

Nous avons donc centré nos efforts sur la réalisation d'une banque enrichie de séquences microsatellites selon Kijas *et al.* (1994) et avons ainsi obtenu 11 marqueurs microsatellites polymorphes (Manuscrit 1). Nous avons également testé les locus microsatellites développés chez *L. migratoria* par Zhang *et al.* (2003) et en avons sélectionné trois (OZC9, OZC35, OZC76). Au total, 14 marqueurs microsatellites ont été disponibles pour les analyses génétiques (voir les manuscrits 3 et 4). Il est important de noter que s'il a été facile d'isoler un grand nombre de séquences microsatellites chez *L. migratoria*, avec un pourcentage de 'vrais' clones positifs dans la banque de 7%, la caractérisation et l'optimisation des marqueurs microsatellites a été difficile. Les principales difficultés concernent l'obtention de fréquents patrons d'amplification multibandes et de déficits forts en génotypes hétérozygotes (Fig. 7). Ces caractéristiques semblent être générales aux orthoptères et lépidoptères (Meglecz *et al.* 2004 ; Zhang *et al.* 2003)

## 2. PATRONS D'AMPLIFICATION MULTI-BANDES

L'observation de patrons d'amplification multibandes a concerné 40% des séquences microsatellites testées (Manuscrit 1). Ce résultat suggère que les séquences microsatellites et notamment leurs régions flanquantes sont fréquemment répétées plusieurs fois dans le génome de *L. migratoria*. Nous avons estimé la proportion de

**Fig. 6. La mitochondrie de *L. migratoria* et les risques de copies nucléaires.**



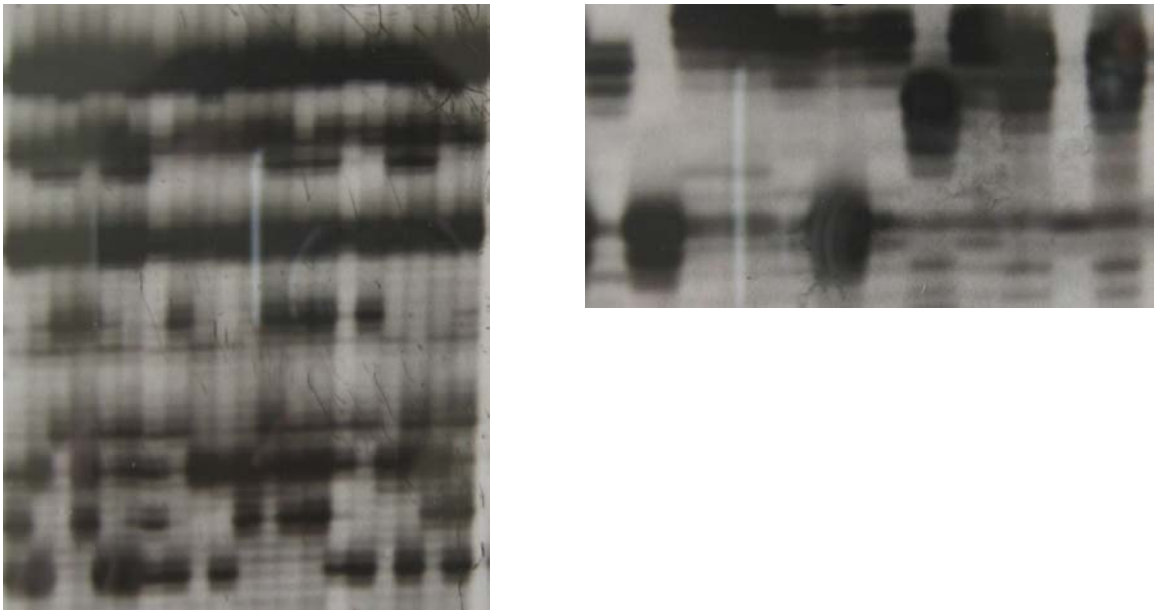

**Copies nucléaires identifiées chez *Locusta migratoria***

**Copies nucléaires identifiées chez d'autres espèces de criquets (*Shistocerca gregaria*, *Podisma pedestris*, *Chortippus parallelus*, *Italopodisma sp.*)**

[1] Gellissen et Michaelis 1987
[2] Vaughan et al. 1999
[3] Bensasson et al. 2000
[4] Zhang et al. 1996

**Fig. 7. Exemples d'un patron d'amplification multibandes (a) et d'un déficit évident en génotypes hétérozygotes (b) aux locus microsatellites de *L. migratoria*.** Les PCR tests sont révélées par marquage radioactif ($P^{33}$). Les locus sélectionnés ont ensuite été génotypés en routine à l'aide d'un séquenceur automatique (marquage par fluorescence des amorces).

**(a)** **(b)**



séquences microsatellites avec un fort degré de similarité, à partir de 135 séquences de *L. migratoria* produites au laboratoire ou récupérées sur Genebank, selon Meglecz *et al.* (2004) (Encadré 1). Cette proportion est élevée (41%) et comparable à celles trouvées chez des espèces de lépidoptères pour lesquels la mise au point de marqueurs microsatellites a posé de grandes difficultés.

Lors de la caractérisation des marqueurs microsatellites, il est possible de réduire significativement le nombre de séquences microsatellites groupées en familles. Par exemple, van't Hof *et al.* (2005) proposent d'identifier des séquences groupées en famille dans un jeu de séquences préliminaire afin de cribler, juste avant l'étape de séquençage des clones positifs, les plasmides par des PCR multiplex amplifiant ces séquences. Alternativement, il est aisé de vérifier, juste avant de dessiner les amorces, que les séquences flanquantes de microsatellites n'ont pas de similarités entre les locus à l'aide du programme BLAST (Altschul *et al.* 1990). Cependant, si seul un nombre limité de clones/séquences sont disponibles, le fait qu'une séquence soit trouvée en un seul exemplaire n'implique pas nécessairement qu'elle soit en copie unique dans le génome.

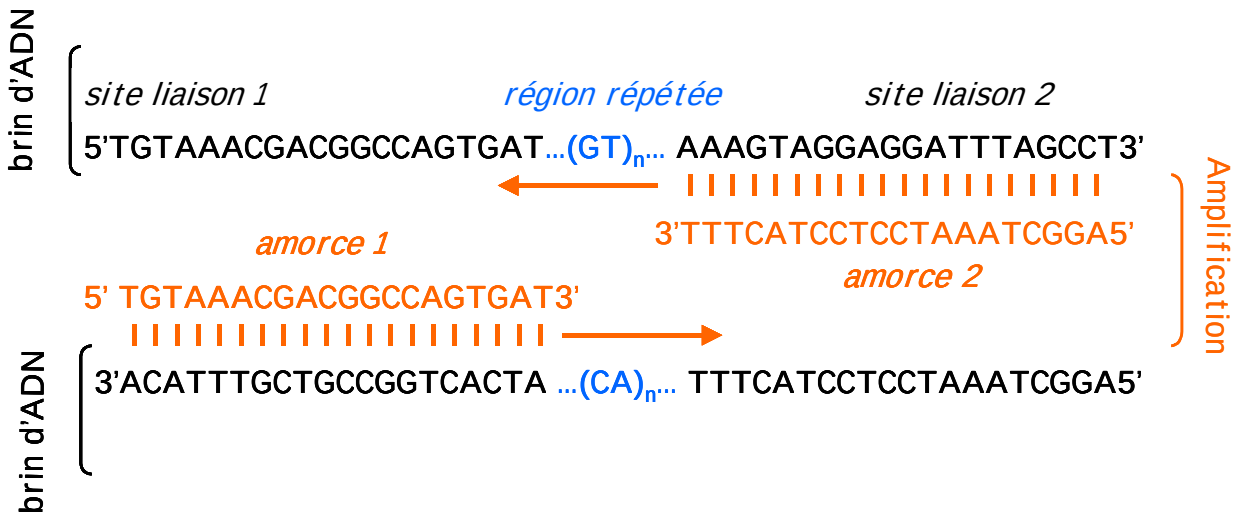
## 3. ALLELES NULS

### 3.1. Prévalence d'allèles nuls et facteurs évolutifs

Les techniques utilisées pour visualiser les variants microsatellitaires (amplification par PCR en utilisant des amorces définies à partir de la séquence d'une copie de gène clonée) ont pour conséquence la présence potentielle d'allèles nuls définis comme des allèles non visibles car non amplifiables par PCR du fait de mutation(s) dans les sites de liaison des amorces (Fig. 8). Nous avons décrit par simulations informatiques les taux et distribution d'allèles nuls aux niveaux intra-populationnels et inter-populationnels, ainsi que les facteurs évolutifs modifiants ces taux (taille de populations, taux de mutations dans les régions flanquantes où sont dessinées les amorces, modèles mutationnels, niveaux de divergence, et flux de gènes entre populations) (Manuscrit 2). Les taux d'allèles nuls sont élevés dans les populations caractérisées par des effectifs importants, une forte instabilité mutationnelle des régions flanquantes et une divergence évolutive forte vis-à-vis de la population à partir de laquelle les séquences d'amorces ont été

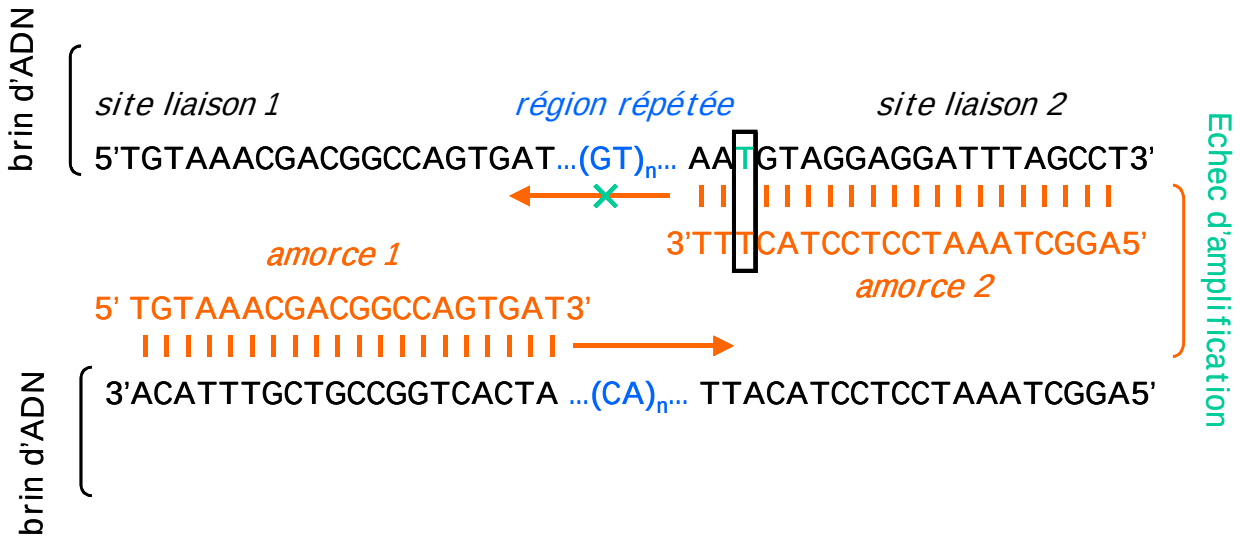
**Fig. 8. Représentation du résultat de l'amplification par Polymerase Chain Reaction (PCR) en absence (a) et en présence (b) d'une variation nucléotidique des régions flanquantes de la région répétée du marqueur microsatellite.**

**(a)**




Le parfait appariement entre les deux amorces, dessinées dans les séquences flanquantes de la région répétée, et le brin d'ADN assure une amplification par PCR du microsatellite. La taille en paires de bases d'une copie de gène, qui reflète premièrement le nombre de motifs répétés, peut alors être révélée par électrophorèse.

**(b)**




La variation nucléotidique dans les sites de liaison (ici substitution de A → T au niveau de l'amorce 2) peut-être responsable d'un mauvais appariement d'une amorce au brin d'ADN, empêchant alors son hybridation durant la PCR. Dans ce cas, le microsatellite n'amplifie pas et l'allèle non détecté (nul) apparaît absent à l'état homozygote et est dominé par les allèles (visibles) avec lesquels il est hétérozygote.

définies. Ces taux dépendent en revanche peu des taux de mutation au niveau de la région répétée et de l'intensité des flux de gènes. Ainsi, il est probable que les grandes tailles de populations ainsi que des taux de mutations forts dans les régions flanquantes soient responsables de la forte fréquence d'allèles nuls observée aux marqueurs microsatellites chez *L. migratoria*.

### 3.2. Optimisation des marqueurs microsatellites enclins aux allèles nuls

Lors de l'optimisation des marqueurs microsatellites, des tests systématiques sur un faible nombre d'individus issus d'une même population (par exemple 8) permettent d'exclure les locus présentant des déficits évidents de génotypes hétérozygotes (*e.g.*, Fig. 7). Cependant, ces tests sommaires peuvent s'avérer insuffisants si la prévalence d'allèles nuls aux marqueurs microsatellites est très élevée dans l'espèce, et ce d'autant plus qu'il existe une structuration géographique de la variation génétique au sein de l'espèce. Dans ce cas, une proportion importante des populations échantillonnées montreront de fortes déviations à l'équilibre d'Hardy-Weinberg (*i.e.*, déficits en génotypes hétérozygotes) pour de nombreux marqueurs microsatellites utilisés suite au test précédent. C'est ce que nous avons observé chez *L. migratoria*. Quarante-sept pour cent des séquences microsatellites ont été exclues à l'issue du test sommaire de déficit d'hétérozygotes (Manuscrit 1). Néanmoins, la majorité des 14 marqueurs microsatellites sélectionnés à la suite de ce test ont montré de fortes déviations à l'équilibre d'Hardy-Weinberg à une majorité des populations génotypées (*i.e.*, déficits en génotypes hétérozygotes ; voir les manuscrits 3 et 4).

Dans ce cas, il convient dans un premier temps de vérifier que les déficits en génotypes hétérozygotes sont effectivement liés à la présence d'allèles nuls et non à un écart à la panmixie ou à des erreurs de génotypage, tels que l'amplification préférentielle des allèles courts (Gagneux *et al.* 1997) ou le glissement de la Taq durant l'amplification (Shinde *et al.* 2003). Un protocole de laboratoire rigoureux (*e.g.*, vérifier la qualité et la quantité des échantillons d'ADN, amplifier plusieurs fois chaque échantillon d'ADN, etc) ainsi que l'utilisation du programme MICRO-CHECKER (van Oosterhout *et al.* 2004), qui discrimine statistiquement entre non-panmixie, erreurs de génotypage et allèles nuls comme causes potentielles des déficits en hétérozygotes observés, permettent de confirmer la présence d'allèles nuls comme la cause la plus probable des déficits observés. L'application de cette méthodologie à nos données

microsatellites de *L. migratoria* analysées a montré que les déficits en génotypes hétérozygotes observés sont très probablement dus à la présence d'allèles nuls (voir les Manuscrits 3 et 4). La fréquence moyenne d'allèles nuls, estimée selon l'algorithme de Dempster *et al.* (1977) sur l'ensemble des populations et des locus génotypés, est égale à 0.19 (Manuscrit 3).

### 3.3. Analyse de jeux de données en présence d'allèles nuls

La présence d'allèles nuls est susceptible de biaiser les estimateurs de diversité et de différenciation génétique. L'étude de fichiers de données microsatellites simulés, en présence et en absence d'allèles nuls, nous a permis de mesurer les biais causés par la présence d'allèles nuls sur les analyses de génétique des populations les plus couramment utilisées (Manuscrit 2). Les résultats principaux montrent que la présence d'allèles nuls a pour conséquence une sur-estimation du niveau de différentiation génétique entre populations, mesuré à l'aide du $F_{ST}$ et des distances génétiques, et une sous-estimation du niveau de diversité génétique à l'intérieur des populations, mesuré à l'aide des hétérozygoties observée et dans une moindre mesure attendue, du nombre d'allèles corrigé ou non par la taille d'échantillonnage, et de la variance de taille allèlique (voir la Fig. 9 pour une illustration sur nos données expérimentales). Ces analyses nous ont également permis de formuler des recommandations pour les écologistes moléculaires travaillant sur des fichiers de génotypes microsatellites présentant un taux d'allèles nuls conséquent :

(1) utiliser l'hétérozygotie attendue (Nei 1987) pour estimer la diversité génétique à l'intérieur des populations, dans la mesure où le biais négatif du aux allèles nuls est faible pour cette statistique (Manuscrit 4);

(2) estimer la différenciation génétique entre les populations avec l'estimateur non biaisé $F_{ST}^{\{ENA\}}$. L'estimation consiste à (i) estimer la fréquence d'allèles nuls et d'allèles visibles selon l'algorithme de Dempster *et al.* (1977), (ii) allouer un seul état aux allèles nuls, non présent dans le fichier original, (iii) ajuster les fréquences génotypiques en se basant sur les estimations (i) des fréquences allèliques, et (iv) calculer le $F_{ST}$ selon Weir (1996) en restreignant le calcul aux états allèliques visibles. Cet algorithme est implémenté dans le programme FreeNA téléchargeable à l'adresse http://www.montpellier.inra.fr/URLB/ (Manuscrit 2) ;

**Fig. 9. Illustration des effets des allèles nuls sur la variation génétique inter-populations (exemple du $F_{ST}$) et intra-populations (exemple du nombre d'allèles) sur les données génotypiques à 14 marqueurs microsatellites de 8 populations d'Europe de l'Ouest de *L. migratoria*.** Les locus se classent, en fonction de leur fréquence d'allèles nuls moyennée sur l'ensemble des populations ($r$), de la manière suivante : 1 locus ne présente pas d'allèles nuls (équilibre d'Hardy-Weinberg et $r < 0.05$), 7 locus ont une fréquence d'allèles nuls moyenne ($0.05 \leq r < 0.20$) et 6 locus ont une fréquence d'allèles nuls forte ($r \geq 0.20$). La fréquence d'allèles nuls $r$ est de 0.20 lorsque elle est moyennée sur l'ensemble des locus.

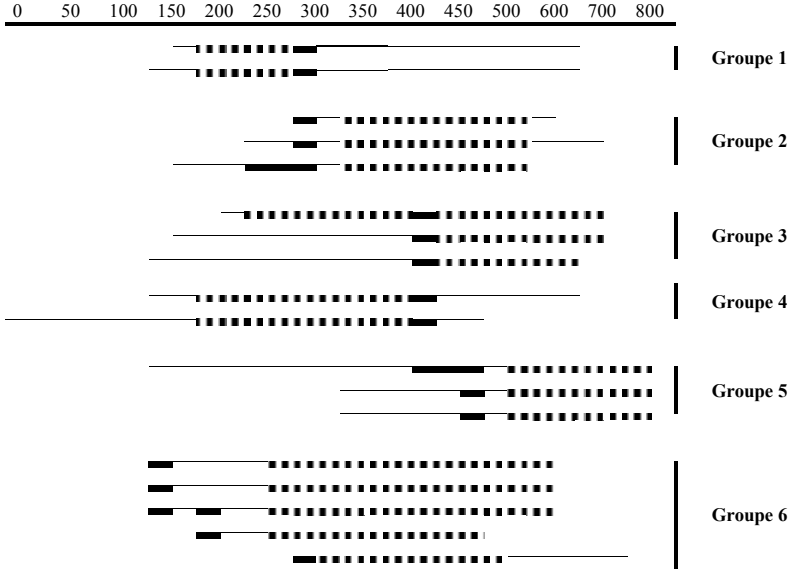
| | $F_{ST}$ global | Nombre d'allèles moyen |
|---|---|---|
| 7 locus avec $0.05 \leq r < 0.20$ | 0.041 | 13.7 |
| les 14 locus ensemble ($r$ moyen = 0.20) | 0.051 | 13.3 |
| 6 locus avec $r \geq 0.20$ | 0.072 | 11.9 |

(3) pour la construction de dendogrammes, utiliser la distance génétique de Cavalli-Sforza et Edwards' (1967) sans correction particulière, puisque cette distance génétique est peu affectée et de manière relativement similaire par les allèles nuls pour une large gamme de temps de divergence entre les populations. Par conséquent, la présence d'allèles nuls biaise peu la topologie des arbres construits à partir de cette distance et de l'algorithme Neighbor-Joining (Saitou et Nei 1987) (Manuscrits 2 et 3) ;
(4) utiliser sans précaution particulière l'analyse BAPS de Corander *et al.* (2003 ; 2004) qui détermine les principaux ensembles populationnels (*i.e.*, en regroupant les populations panmictiques qui échangent des migrants à un taux élevé). On notera que BAPS discrimine mieux les différents ensembles génétiques en présence qu'en absence d'allèles nuls. Cependant, sa puissance reste faible pour des niveaux de flux de gènes élevés (Manuscrit 4).

**ENCADRE 1 –** SIMILARITES ENTRE REGIONS FLANQUANTES DE MICROSATELLITES CHEZ *L. MIGRATORIA*

Les séquences de 136 microsatellites contenant une région d'au moins trois motifs répétés ininterrompus ont été obtenues, à partir d'une banque partielle d'ADN nucléaire réalisée au CBGP (127 séquences) et des entrées Genbank (8 séquences ; Zhang *et al.* 2003). A l'aide du programme BLAST (Altschul *et al.* 1990), nous avons comparé toutes les séquences paire par paire. Nous avons fixé le nombre d'appariements attendu par chance seulement à $10^{-10}$ et ignoré systématiquement les régions de motifs répétés en filtrant les séquences à faible complexité (filtre DUST ; Hancock et Armstrong 1994). Cinquante-six séquences microsatellites (soit 41%) ont ainsi été groupées en treize familles (voir la Fig. 1 pour les groupes les plus homologues). Chaque famille comprend de 2 à 23 séquences microsatellites. Les régions similaires sont longues (de 50 à 300 paires de bases) et ont des identités fortes (> 85%). Elles ne concernent jamais la totalité de la séquence clonée, ce qui exclut la possibilité de séquences redondantes. L'état de groupement des séquences microsatellites n'est pas significativement dépendant ni de la longueur du motif répété (Test de Mann-Whitney; $P = 0.133$) ni de la complexité de la région de motifs répétés (*i.e.* microsatellites simples *versus* microsatellites interrompus ou composés ; Test exact de Fisher ; $P = 1$). Nous avons également testé si l'occurrence de familles de séquences était associée à la présence de motifs répétés courts spécifiques des microsatellites en comparant paire par paire 17 séquences ne contenant pas de région de motifs répétés courts avec elles-mêmes et avec le fichier de séquences contenant une région de motifs répétés. Aucune de ces séquences ne s'est appariée. L'occurrence de familles de séquences semble ainsi être associée à la présence de motifs répétés courts spécifiques des microsatellites (Test exact de Fisher ; $P = 0.008$). La proportion de séquences microsatellites groupées chez *L. migratoria* (41%) est marginalement ou non différente de celles des séquences microsatellites des espèces de lépidoptères *Parnassius apollo* et *Euphydryas aurinia* étudiées par Meglecz *et al.* (2004 ; Tests exacts de Fisher : $P = 0.05$ et $P = 1$, respectivement). En revanche, elle est significativement supérieure à celle de l'hyménoptère *Apis mellifera*, pour lequel la mise au point de marqueurs microsatellites, clonés selon le même protocole que *L. migratoria*, ne pose pas de difficultés particulières (Test exact de Fisher ; $P < 10^{-4}$ ). On notera enfin que la proportion de séquences microsatellites groupées chez *L. migratoria* (41%) peut à elle seule expliquer la forte proportion de séquences microsatellites que nous avons exclues pour raison de patrons d'amplification multibandes lors de la réalisation de la banque microsatellite (voir manuscrit 1 ; Test exact de Fisher ; $P = 1$).

**Fig. 1.** Exemple de six groupes de séquences microsatellites présentant de très fortes similitudes ( $\varepsilon < 10^{-40}$)




Le nombre d'appariements attendu par chance seulement ( $\varepsilon$ ) est inférieur à $10^{-40}$ pour l'ensemble des paires de séquences microsatellites présentées dans cette figure. Les six groupes de séquences microsatellites représentent 13% des séquences analysées. Les lignes continues épaisses représentent les régions de motifs répétés courts, les lignes continues fines les régions non répertoriées comme uniques dans l'échantillon de séquences étudiées, et les lignes discontinues épaisses les régions similaires aux autres régions du même groupe.

## MANUSCRIT 1 – CHARACTERIZATION AND PCR MULTIPLEXING OF POLYMORPHIC MICROSATELLITE LOCI FOR THE LOCUST *LOCUSTA MIGRATORIA*

Marie-Pierre Chapuis, Anne Loiseau, Yannis Michalakis, Michel Lecoq
et Arnaud Estoup

Molecular Ecology Notes 2005 – **5**: 554-557.

Molecular Ecology Notes (2005) **5**, 554–557 doi: 10.1111/j.1471-8286.2005.00987.x

PRIMER NOTE

# Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*

MARIE-PIERRE CHAPUIS,* ANNE LOISEAU,* YANNIS MICHALAKIS,† MICHEL LECOQ‡ and ARNAUD ESTOUP*

**Centre de Biologie et de Gestion des Populations, INRA, Campus International de Baillarguet CS 30016, 34988 Montferrier/Lez, France, †Centre d'Etudes sur le Polymorphisme des Microorganismes, UMR CNRS-IRD 9926, Montpellier cedex 5, France, ‡Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Campus International de Baillarguet 34398 Montpellier cedex 5, France*

**Abstract**

**Because of the scarcity of polymorphic genetic markers available in locust species, only a few population genetics studies have been carried out on this taxon. We isolated and characterized 11 polymorphic microsatellite loci in the pest locust *Locusta migratoria capito*, and described experimental conditions for polymerase chain reaction (PCR) multiplexing and simultaneously genotyping these loci. The number of alleles per locus ranged from six to 25, and the expected heterozygosity ranged from 0.431 to 0.957. Results of cross-taxon amplification tests are reported in six other *Locusta migratoria* subspecies, six species of the Oedipodinae subfamily and two other pest locust species.**

*Keywords*: locust, *Locusta migratoria*, microsatellites, null alleles, Orthoptera, pest

*Received 4 February 2005; revision accepted 25 February 2005*

The migratory locust, *Locusta migratoria*, is an agricultural pest of major importance in large areas of the Ancient world. The species displays unusual outbreaks with an irregular cycle and swarming individuals typical of a high-density distinctive form, the phase *gregaria* (Uvarov 1966). These swarms are increasing either as a result of particular weather conditions or of anthropogenic changes of the environment (e.g. Lecoq & Sukirno 1999). Nine geographical taxa of *L. migratoria* with different propension to display outbreaks were described and considered as subspecies (review in COPR 1982). Although the phenotypic study of Farrow & Colless (1980) casts doubt on the validity of this classification, most of the studies focus on outbreaking subspecies (generally *Locusta migratoria migratorioides*). Because of the scarcity of polymorphic genetic markers available in locust species, only a few population genetics studies have been achieved in this taxon to better understand swarming behaviour and taxonomy. Population genetic analysis would allow measuring gene flow between *L. migratoria* subspecies and shed light on the swarming dynamics in swarming subspecies. Moreover, a high level of polymorphism is necessary to address these questions as the migratory locust has a great dispersal ability which probably generates substantial gene flow between populations over large geographical areas. Here we present the development of 11 polymorphic microsatellite loci for *L. migratoria*.

A partial DNA library of 2640 recombinant clones was obtained from a single *Locus migratoria capito* individual originating from Madagascar, following the enrichment protocol from Kijas *et al*. (1994). The library was screened with $(GT)_{10}$ and $(CT)_{10}$ probes following Estoup *et al*. (1998) (detailed protocols available at http://www.inapg.inra.fr/dsa/microsat/microsat.htm). Two hundred and forty-one of the recombinant clones (i.e. 9%) showed a strong hybridization signal. Among them, 160 inserts were sequenced using the BigDye technology and an ABI 3730xl automated sequencer (Applied Biosystems). One hundred and twenty-nine of them contained a microsatellite sequence and 53 were tested for polymerase chain reaction (PCR) amplification using fluorescent primers and a MegaBACE DNA sequencer (Amersham Biosciences). Individual DNA extractions were performed from a 2-mm long piece of leg following a standard CTAB-based protocol (Doyle & Doyle 1987). PCR

Correspondence: Arnaud Estoup, Fax: +33 (0)499623345; E-mail: estoup@ensam.inra.fr

**Table 1** Microsatellite loci developed for the locust *Locusta migratoria capito*

| Locus | Core repeat (cloned allele) | Size (bp) | Size range (bp) | $N_a$ | $f$ | $H_O$ | $H_E$ | HW test | Primer (μM) | $T_{ann}$ (°C) | Primer sequence (5′–3′) | GenBank Accession no. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PCR multiplex set 1 | | | | | | | | | | | | |
| *LM1-88* | $(AC)_{12}$ | 106 | 102–118 | 9 | 0.362 | 0.759 | 0.781 | | 0.2 | 60 | F: NED-GATGGGATACCGACTAGAGC<br>R: CTCAAGCCTACCCTCTCAAG | AY883024 |
| *LM10-78* | $(CT)_{16}$ | 112 | 98–141 | 25 | 0.121 | 0.897 | 0.957 | | 0.2 | 60 | F: HEX-CCACTTCTGTCAATGCACTC<br>R: CACTTGCAACGGGTTCTAC | AY883025 |
| *LM2-B* | $(GT)_{24}$ | 337 | 309–347 | 13 | 0.192 | 0.423 | 0.896 | * | 0.3 | 60 | F: FAM-AGACTGGCAGTGGAGAGGAC<br>R: GACGGCAATTATAAGGCGAC | AY883026 |
| *LM2-Λ* | $(GT)_{18}$ | 245 | 233–272 | 13 | 0.135 | 0.615 | 0.926 | * | 0.2 | 60 | F: HEX-GCTTCACTTTGTCGTTCTAGG<br>R: GCGTAAAATCCTCAGCCTGC | AY883027 |
| *LM3-Ω* | $(GT)_{16}$ | 202 | 174–211 | 16 | 0.259 | 0.759 | 0.891 | | 0.2 | 60 | F: NED-CTATTCAGGGCACTACTTTGAG<br>R: CAGTGGTCATCTATCCCTAAAC | AY883028 |
| *LMT-113* | $(CA)_{15}$ | 297 | 294–329 | 19 | 0.224 | 0.828 | 0.913 | | 0.3 | 60 | F: NED-CAGAATTTCAGACACTACCTTCC<br>R: CGTCTCGTAAAGTAATGTGCG | AY883029 |
| *LMT-137* | $(AC)_4$ $AA(CA)_{22}$ | 152 | 113–188 | 14 | 0.286 | 0.889 | 0.884 | | 0.2 | 60 | F: FAM-CTTGAACTAACTTAATGCTAAGG<br>R: ATATGTAGCAGCGTAGAACTTG | AY883030 |
| *LMT-177* | $(GT)_9$ $TT(GA)_5$ | 205 | 205–217 | 6 | 0.741 | 0.310 | 0.431 | | 0.3 | 60 | F: HEX-GCCTATTTGAGACAGTGCCT<br>R: CTGATTCAATCTCATAACCTGG | AY883031 |
| PCR multiplex set 2 | | | | | | | | | | | | |
| *LM10-180* | $(CA)_{11}$ $AA(CA)_7$ | 266 | 251–280 | 10 | 0.379 | 0.379 | 0.791 | * | 0.2 | 57 | F: HEX-TCTGACCTTCCTGAATATGAG<br>R: GAGTTCGGTCATGCATTG | AY883032 |
| *LM11-121* | $(CA)_{15}$ | 360 | 362–396 | 20 | 0.185 | 0.926 | 0.933 | | 0.5 | 57 | F: FAM-GATGTTGTAAGTGTCCGCCA<br>R: GCAATATTTCGATAAGGTAGCAC | AY883033 |
| *LMT-133* | $(CA)_{15}$ | 204 | 204–226 | 8 | 0.558 | 0.423 | 0.652 | * | 0.2 | 57 | F: HEX-TCACCATTGAATGTCCTATACAG<br>R: GATGAAACTAACTTTTGAACTGG | AY883034 |

The observed size range, number of alleles ($N_a$), frequency of the most common allele ($f$), and observed ($H_O$) and expected ($H_E$) heterozygosities were estimated from 30 individuals collected in a population from Madagascar. HW test: Hardy–Weinberg exact test (Raymond & Rousset 1995), *: significant probability test with $\alpha = 0.05$ using the sequential Bonferroni correction (Rice 1989). Primer: primer concentration for both forward and reverse primers, $T_{ann}$: PCR annealing temperature.

amplifications were performed using a Mastercycler thermocycler and consisted of an initial denaturing step of 15 min at 94 °C; 35 cycles of 30 s at 94 °C, annealing for 90 s (see Table 1 for annealing temperatures), and extension at 72 °C for 60 s; and a 10 or 30 min final extension step. Eleven polymorphic loci with a low number of aspecific and/or stutter bands (21 out of 53 loci excluded based on this criterion) and no obvious deficit in heterozygous genotypes in a sample of eight individuals from Madagascar (15 out of 32 remaining loci excluded) were selected for further population studies. Primer sequences and PCR conditions are given for each selected locus in Table 1. The sequences of the 11 microsatellite loci have been deposited in the GenBank database (accession numbers: AY883024 to AY883034). According to annealing temperatures, two sets of eight and three loci were independently coamplified using a multilocus amplification kit (QIAGEN) in a 10-μL volume containing 1× QIAGEN Multiplex Master Mix (+Q for locus set 1 and –Q for set 2; Table 1), 2–5 μM of each primer and 2 μL of genomic DNA (≤ 1 μg DNA). The fluorescent dye (FAM, HEX or NED) of the forward primer of each microsatellite was chosen in order to allow detecting alleles at all 11 loci in a single electrophoresis run (Table 1).

The level of polymorphism and allelic distribution of the 11 microsatellite loci were estimated using GENEPOP version 3.4 (Raymond & Rousset 1995) by typing 30 individuals collected in a population from Madagascar (*L. m. capito*). The number of alleles per locus and the expected heterozygosities ranged between 6 and 25, and 0.431 and 0.957, respectively (Table 1). Although the 11 loci were preselected in order to limit the risk of null alleles, large differences between observed and expected heterozygosities were observed at some loci, and are likely to be due to the presence of null alleles at those loci (Table 1). A substantial proportion of *L. migratoria* microsatellites with heterozygote deficit was also observed in Zhang *et al*. (2003), suggesting that null alleles may be particularly frequent in this species.

All microsatellites developed from *L. m. capito* were successfully analysed using the previously discussed PCR conditions in six other *L. migratoria* subspecies, giving allelic patterns of similar quality and high levels of polymorphism (Table 2). The conservation of the primer

**Table 2** Cross-species PCR tests for microsatellite loci developed from *Locusta migratoria capito* in six *Locusta migratoria* subspecies, six species of the Oedipodinae subfamily and two pest species of the Cyrtacanthacridinae subfamily

| | *LM1-88* | *LM10-78* | *LM2-B* | *LM2-Λ* | *LM3-Ω* | *LMT-113* | *LMT-137* | *LMT-177* | *LM10-180* | *LM11-121* | *LMT-133* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Species *Locusta migratoria* | | | | | | | | | | | |
| Arabian subspecies | 4 | 11 | 10 | 11 | 11 | 9 | 12 | 7 | 9 | 12 | 6 |
| (*N*=8) | 96–147 | 107–149 | 313–343 | 241–293 | 178–213 | 293–333 | 123–186 | 205–217 | 254–289 | 358–384 | 206–222 |
| Australian subspecies | 6 | 9 | 6 | 5 | 6 | 13 | 11 | 2 | 5 | 11 | 6 |
| (*N*=8) | 96–116 | 109–129 | 303–335 | 245–272 | 189–204 | 297–342 | 119–188 | 206–217 | 250–275 | 351–380 | 200–218 |
| *L. m. cinerascens* | 7 | 9 | 9 | 4 | 10 | 4 | 11 | 6 | 5 | 8 | 8 |
| (*N*=8) | 88–106 | 107–129 | 311–339 | 245–264 | 176–210 | 297–324 | 113–184 | 203–215 | 259–277 | 359–382 | 208–226 |
| *L. m. manilensis* | 6 | 8 | 6 | 6 | 7 | 10 | 10 | 5 | 8 | 10 | 5 |
| (*N*=8) | 96–126 | 107–133 | 303–335 | 249–274 | 185–211 | 303–337 | 119–196 | 206–215 | 252–273 | 342–380 | 200–218 |
| *L. m. migratoria* | 9 | 6 | 8 | 4 | 9 | 3 | 12 | 6 | 6 | 11 | 6 |
| (*N*=8) | 96–166 | 105–133 | 313–347 | 258–266 | 170–209 | 298–308 | 113–200 | 203–215 | 258–279 | 357–388 | 190–218 |
| *L. m. migratorioides* | 6 | 14 | 9 | 8 | 13 | 8 | 13 | 9 | 6 | 12 | 5 |
| (*N*=8) | 88–110 | 104–127 | 311–345 | 245–270 | 166–210 | 305–329 | 123–188 | 203–215 | 252–272 | 345–380 | 202–220 |
| Subfamily Oedipodinae | | | | | | | | | | | |
| *Gastrimargus africanus* | — | — | — | — | — | — | — | — | 3 (?) | — | — |
| (*N*=4) | | | | | | | | | 255–267 | | |
| *Oedaleus decorus* | 4 (?) | — | — | — | — | — | — | — | — | — | — |
| (*N*=3) | 96–104 | | | | | | | | | | |
| *Oedaleus senegalensis* | — | — | — | — | — | — | — | — | — | — | — |
| (*N*=3) | | | | | | | | | | | |
| *Oedaleus virgula* | — | — | 3 (?) | — | — | — | — | 2 | — | — | — |
| (*N*=4) | | | 309–315 | | | | | 206–207 | | | |
| *Oedipoda germanica* | — | — | — | — | — | — | — | — | — | — | — |
| (*N*=5) | | | | | | | | | | | |
| *Sphingonotus caerulans* | — | — | — | — | — | — | — | — | — | — | — |
| (*N*=3) | | | | | | | | | | | |
| Subfamily Cyrtacanthacridinae (pest species) | | | | | | | | | | | |
| *Nomadacris septemfasciata* (*N*=4) | — | — | — | — | — | — | — | — | — | — | — |
| *Schistocerca gregaria* | — | — | — | — | — | — | — | — | — | — | — |
| (*N*=4) | | | | | | | | | | | |

The number of alleles and the allelic size range are based on *N* individuals. Amplification failure is indicated by a dash. (?): specific PCR product but uncertainty on the allelic pattern.

sequences and the level of polymorphism of the selected loci were also examined in eight other grasshopper species of the family Acrididae. Six species are of the subfamily Oedipodinae (*Gastrimargus africanus*, *Oedaleus decorus*, *Oedaleus senegalensis*, *Oedaleus virgula*, *Oedipoda germanica* and *Sphingonotus caerulans*). Those species represent the closest related taxa of the monospecific genus *Locusta*. Two other pest locusts, the desert locust *Schistocerca gregaria* and the red locust *Nomadacris septemfasciata* (subfamily Cyrtacanthacridinae) were also tested. For each species, extraction and PCR amplification were performed following the procedure described previously and repeated twice. Results reported in Table 2 show that these loci will not be useful in other locust or grasshopper species.

Including the set of primers developed by Zhang *et al*. (2003), 19 microsatellite markers are now available for population genetic analysis in *L. migratoria*.

## Acknowledgements

We are grateful to Sekouba Foumavro, Alex Franc, Yayat Hidayat, David Hunter, Maïga Idrissa, Adnan Khan and Oleg N. Naumovich for providing us with specimens of some studied species. This work was partly funded by the scientific SPE department of INRA.

## References

COPR (1982) *The Locust and Grasshopper Agricultural Manual*. Centre for Overseas Pest Research, London.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, **19**, 11–15.

Estoup A, Gharbi K, SanCristobal M *et al*. (1998) Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 715–723.

Farrow RA, Colless DH (1980) Analysis of the interrelationships of geographical races of *Locusta migratoria* (Linnaeus) (Orthoptera:

Acrididae) by numerical taxonomy, with special reference to sub-speciation in the tropics and affinities of the Australian race. *Acrida*, **9**, 77–99.

Kijas JMH, Fowler JCS, Garbet CA, Thomas MR (1994) Enrichment of microsatellites from citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particules. *BioTechniques*, **16**, 657–662.

Lecoq M, Sukirno (1999) Drought and exceptional outbreak of the oriental migratory locust, *Locusta migratoria manilensis* (Meyen 1835) in Indonesia (Orthoptera: Acrididae). *Journal of Orthoptera Research*, **8**, 153–161.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Uvarov BP (1966) *Grasshoppers and Locusts*. *A Handbook of General Acridology*. Volume 1. Cambridge University Press, UK.

Zhang DX, Yan LN *et al*. (2003) Microsatellite DNA markers in the migratory locust, *Locusta migratoria*: isolation, variability, cross-species amplification and mode of inheritance. *Molecular Ecology Notes*, **3**, 483–486.

## MANUSCRIT 2 – MICROSATELLITE NULL ALLELES AND ESTIMATION OF POPULATION DIFFERENTIATION

Marie-Pierre Chapuis et Arnaud Estoup

Molecular Biology and Evolution 2006 – 10.1093/molbev/msl191

**MBE Advance Access published December 5, 2006**

# MICROSATELLITE NULL ALLELES AND ESTIMATION OF POPULATION DIFFERENTIATION

**Research Article**

Marie-Pierre Chapuis[1,2,3] and Arnaud Estoup[1]

[1]Centre de Biologie et de Gestion des Populations, Institut National pour la Recherche Agronomique, Campus International de Baillarguet CS 30016, 34988 Montferrier / Lez, France

[2]Génétique et Evolution des Maladies Infectieuses, UMR 274 CNRS-IRD, 911 avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France

[3]Centre de coopération internationale en recherche agronomique pour le développement, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

**Correspondence**:

Marie-Pierre CHAPUIS
Centre de Biologie et de Gestion des Populations
Institut National de la Recherche Agronomique
Campus International de Baillarguet CS 30 016
34988 Montferrier / Lez, FRANCE
Phone: +33 (0)4.99.62.33.31
Fax: +33 (0)4.99.62.33.45
E-mail: chapuimp@ensam.inra.fr

**Keywords**: coalescent, microsatellite, null alleles, population differentiation, *F*-statistics, genetic distances

**Running head:** microsatellite null alleles

**Nonstandard abbreviations:**

*R*: repeat region of a microsatellite locus

*B*: sequence corresponding to the binding sites of the two microsatellite locus primers

*VA* dataset: simulated dataset that includes all alleles

*NA* dataset: simulated dataset (the same individuals) without the null alleles

*CNA* dataset: the *NA* simulated dataset after correction

$\hat{r}_C$, $\hat{r}_B$, $\hat{r}_D$: null allele frequency estimators of Chakraborty et al. (1992), Brookfield (1996), and Dempster, Laird, and Rubin (1977), respectively

*INA*: calculation of the differentiation measures ($F_{ST}$ and genetic distance) from the *CNA* dataset when the null allele size is included

*ENA*: calculation of the $F_{ST}$ from the *CNA* dataset when the null allele size is excluded

**Abstract**

Microsatellite null alleles are commonly encountered in population genetics studies, yet little is known about their impact on the estimation of population differentiation. Computer simulations based on the coalescent were used to investigate the evolutionary dynamics of null alleles, their impact on $F_{ST}$ and genetic distances, and the efficiency of estimators of null allele frequency. Further, we explored how the existing method for correcting genotype data for null alleles performed in estimating $F_{ST}$ and genetic distances and we compared this method to a new method proposed here (for $F_{ST}$ only). Null alleles were likely to be encountered in populations with a large effective size, with an unusually high mutation rate in the flanking regions, and that have diverged from the population from which the cloned allele state was drawn and the primers designed. When populations were significantly differentiated, $F_{ST}$ and genetic distances were overestimated in the presence of null alleles. Frequency of null alleles was estimated precisely with the algorithm presented in Dempster, Laird, and Rubin (1977). The conventional method for correcting genotype data for null alleles did not provide an accurate estimate of $F_{ST}$ and genetic distances. However, the use of the Cavalli-Sforza and Edwards' (1967) genetic distance corrected by the conventional method gave better estimates than those obtained without correction. $F_{ST}$ estimation from corrected genotype frequencies performed well when restricted to visible allele sizes. Both the proposed method and the traditional correction method have been implemented in a program that is available free of charge at http://www.montpellier.inra.fr/URLB/. We used two published microsatellite datasets based on original and redesigned pairs of primers to empirically confirm our simulation results.

## Introduction

Microsatellites are popular and versatile molecular markers for addressing questions in population genetics and evolution (Estoup and Angers 1998). Observed microsatellite alleles are DNA fragments of different sizes detected by initial amplification using polymerase chain reaction (PCR) and visualization via electrophoresis. Size polymorphism reflects variation in the number of repeats of a simple DNA sequence (2-6 bases long). However, sequencing studies indicate that changes in flanking region sequences also occur at a non-negligible rate (e.g., Angers and Bernatchez 1997; Grimaldi and Crouau-Roy 1997). Such variation in the nucleotide sequences of flanking regions may prevent the primer annealing to template DNA during amplification of the microsatellite locus by PCR, resulting in a null allele. The molecular origin of null alleles (substitution and indel mutations) resulting from polymorphism in the annealing region has been assessed directly by sequencing the annealing sites of microsatellite locus primers for both null and visible alleles (Callen et al. 1993). Other possible causes of microsatellite null alleles include the preferential amplification of short alleles (due to inconsistent DNA template quality or quantity) or slippage during PCR amplification (Gagneux, Boesch, and Woodruff 1997; Shinde et al. 2003). These technical problems associated with amplification will not be considered here.

The presence of microsatellite null alleles has been reported frequently in PCR primer characterization and in population genetics studies (Dakin and Avise 2004). Although microsatellite null alleles have been found in a wide range of taxa, some taxa have a particularly high frequency of null alleles; examples include insects (Lepidoptera, reviewed in Meglecz et al. 2004; Diptera, Lehmann et al. 1997; Orthoptera, Chapuis et al. 2005) and molluscs (Li et al. 2003; Astanei et al. 2005). Interestingly, these are species with large effective population sizes. The association between the presence of null alleles and highly variable flanking regions has been demonstrated repeatedly in molecular studies, and several

studies have suggested that the sequences flanking microsatellites may be less stable than those in other genomic regions (Angers and Bernatchez 1997; Grimaldi and Crouau-Roy 1997; Meglecz et al. 2004). On the other hand, no correlation has been found between null allele frequency and microsatellite unit-repeat length or motif complexity (Li et al. 2003), two factors related to the mutation rate of the microsatellite repeat region (Jin et al. 1996; Chakraborty et al. 1997). The null allele frequency in a congeneric species has been shown to rapidly increase with increasing phylogenetic distance from a focal species (e.g., in the oyster *Crassostrea*; Li et al. 2003). Despite the known prevalence of null alleles, the evolutionary dynamics and patterns of variation of these alleles in populations has never been examined analytically or by computer simulation.

Ninety percent of articles reporting microsatellite loci with null alleles include these loci in their analyses without correction for potential bias (reviewed in Dakin and Avise 2004). Yet null alleles may affect the estimation of population differentiation, for instance, by reducing the genetic diversity within populations (e.g., Paetkau and Strobeck 1995). Markedly, $F_{ST}$ and genetic distances values generally increase with decreasing within-population genetic diversity (Slatkin 1995; Paetkau et al. 1997). The extent to which null alleles may overestimate population differentiation has never been investigated.

Null alleles can be detected in population studies by carefully testing for Hardy-Weinberg (HW) proportions, provided that observed heterozygote deficiencies have no other origin (e.g., Wahlund effect). Various null allele frequency estimators ($\hat{r}$) making use of this property have been developed (Dempster, Laird, and Rubin 1977; Chakraborty et al. 1992; Brookfield 1996). Some authors have attempted to correct for null alleles in population genetic studies by statistical adjustment of the visible allele and genotype frequencies, based on $\hat{r}$ and assuming a single new null allele size common to all genotyped populations (Roques, Duchesne, and Bernatchez 1999). However, experimental studies using various

amplifications (null and non null) to determine the null allele sizes have suggested that null alleles often correspond to alleles with different sizes, and that alleles with the same size may correspond to both null and visible states (Callen et al. 1993; Paetkau and Strobeck 1995; Lehmann, Hawley, and Collins 1996). The efficiency of the null allele frequency estimators and the existing correcting method has not been assessed.

We used computer simulations based on the coalescent (Hudson 1990) to investigate the prevalence and distribution of null allele sizes at microsatellite loci. We then assessed the impact of such null alleles on two statistics traditionally used to estimate population differentiation, $F_{ST}$ and genetic distance. We evaluated the available methods for estimating null allele frequency and population differentiation from datasets with null alleles and propose a new method for estimating $F_{ST}$ in the presence of null alleles. We illustrate our simulation results by verifying empirically the presence and impact of null alleles in two published microsatellite datasets based on original and redesigned pairs of primers (Paetkau and Strobeck 1995; Lehmann, Hawley, and Collins 1996).

## Materials and Methods

### Simulation method

We used a three-step simulation approach described schematically in Figure 1.

**Step 1.** Genotypic data were simulated from an algorithm based on the coalescent (Leblois, Estoup, and Rousset 2003; Paetkau et al. 2004). Two population models were assumed: a migration model and a split population model. In the migration model, two populations of equal effective size $N_e$ exchange migrants at a rate $m$. In the population split model, an ancestral population of $N_e$ individuals splits into two populations, each with the same effective size $N_e$; these two populations then do not exchange any genes for $t$ generations. After the coalescent tree was constructed, we simulated mutational events on this

tree, both within the repeat region of the microsatellite locus (hereafter referred to as *R*; mutation rate $\mu_R$) and in the bases flanking the microsatellite locus for which a mutation is likely to prevent primer binding (hereafter referred to as *B*; mutation rate $\mu_B$). We chose *B* as the 10 pb binding to the 3' end of each 20 pb-long primer, so that only half of the mutations at the binding sites precluded PCR amplification. *R* and *B* were assumed to be completely linked. This assumption is reasonable because of the short physical distance between these regions (less than 300 bp). The number of mutations in *R* and *B* was simulated along each branch of the tree, according to a Poisson distribution with parameters $L\mu_R$ and $L\mu_B$, respectively, where *L* is the length of the branch in generations. Mutation rates $\mu_R$ were assumed to be equal for all loci. The same assumption was made for the mutation rates $\mu_B$. Mutations in *R* followed a symmetric generalized stepwise mutation model without allele size constraints (GSM; Zhivotovsky, Feldman, and Grishechkin 1997; Estoup, Jarne, and Cornuet 2002). Changes in the number of repeat units followed a geometric distribution with a variance of 0.36 (Estoup et al. 2001). Mutations in *B* followed an infinite allele model (IAM; Kimura and Crow 1964). Once genotypic data had been simulated for both *R* and *B*, we randomly selected a gene copy used for the design of the microsatellite primers from a single focal population. This imitates the work of molecular biologists, who design PCR primers based on the sequence of a single gene copy in a given population. The allele state of the *B* region of the selected gene copy (hereafter referred to as *B*-cloned allele state) corresponded to the state of *B* for which PCR amplification was successful. All other *B* allele states were assumed to preclude PCR amplification. Therefore, any *R* gene copy not associated with the *B*-cloned allele state bore a null allele.

**Step 2.** From a single set of genotypic data, three datasets, composed of 60 genes (or 30 diploid individuals) for each population, were generated simultaneously. In the first, all *B* allele states were assumed to allow PCR amplification, so no null alleles were present (*VA*

dataset). Using the second dataset, the *R* alleles not associated with the *B*-cloned allele state were assumed to be null (*NA* dataset). The simulated *NA* genotype dataset was corrected for null alleles following the approach of all empirical population genetics studies to date (*CNA* dataset; Roques, Duchesne, and Bernatchez 1999). Null and visible allele frequencies were first estimated with the algorithm described in Dempster, Laird, and Rubin (1977) and *Supplementary Material* which performed best of all the null allele frequency estimators tested (see '*Results'* section). Homozygous genotype frequencies were then adjusted. We partitioned apparent homozygous counts $n_{ii}$ into true $n^*_{ii}$ and false $n^*_{i0}$ homozygous counts. The true homozygote frequency is $p^*_{ii} = \left( \frac{n^*_{ii}}{n^*_{ii} + n^*_{i0}} \right) \times \frac{n_{ii}}{n}$ with $n$ the number of individuals. Based on the relationships between true genotype counts and frequencies, we obtained the following estimate for homozygote frequency: $\hat{p}_{ii} = \left( \frac{\hat{p}_i}{\hat{p}_i + 2\hat{r}_D} \right) \times \frac{n_{ii}}{n}$, with $\hat{r}_D$ the estimate of null allele frequency. Finally, all null alleles were given a single arbitrary allele size, not present in the original dataset.

**Step 3.** The available method for estimating population differentiation in the presence of null alleles uses *CNA* genotype datasets and is referred to as *INA* (i.e., Including Null Alleles). The $F_{ST}$-estimate at a given locus is the appropriate combination of allele-based estimates for several alleles (Weir 1996). We hence propose a new correction for estimating $F_{ST}$ in the presence of null alleles, in which $F_{ST}$ is estimated from *CNA* datasets, but the calculation is restricted to visible allele sizes (referred to as *ENA* for Excluding Null Alleles). Note that, in this case, the sums of the frequencies of alleles and genotypes are not adjusted to 1. This approach can not be used in the calculation of genetic distances, however, because genetic distances are expressed in terms of the proportions of similar alleles between and within populations, and so the lowest level of integration for such measures is the locus (i.e., the entire set of visible and null alleles).

Tests on simulated datasets

We generated 10,000 simulated datasets for 35 different couples of values of the mutational parameter $N_e \mu_B$ ($10^{-4}$; $10^{-3}$; $10^{-2}$; $10^{-1}$; and 1) and the populational parameter $N_e m$ (0.01; 0.1; 1; and 10) or $t$ (1,000; 10,000; and 100,000) according to the population model considered. It is worth stressing here that the product $N_e \mu_B$, not $\mu_B$ alone, determines the level of variation in binding sites, and hence the prevalence of null alleles in population gene samples. Preliminary simulations showed that the prevalence and allele size distribution of null alleles remained similar for a large range of $N_e \mu_R$ values (results not shown). We therefore fixed the product $N_e \mu_R$ at 1 for all simulations. This resulted in heterozygosity values spanning a large part of the range of heterozygosity generally observed at microsatellite markers (0.5 to 0.8; Takezaki and Nei 1996).

We first tested observations stemming from molecular studies that null alleles at a microsatellite locus are likely to be encountered in populations with a large effective size and/or an unusually high mutation rate in the flanking regions (i.e., large $N_e \mu_B$ values), and in populations that have diverged from the population from which the cloned allele state was drawn and the primers designed. To do so, we determined the range of values and/or combinations of the parameters $N_e \mu_B$ and $N_e m$ or $t$ (according to the population model considered) favoring the presence of null alleles in population gene samples by simulating single-locus *NA* datasets. The simulated loci were categorized, separately for the focal and non-focal population, into three classes of null allele frequency: negligible ($r < 0.05$), moderate ($0.05 \leq r < 0.20$) or large ($r \geq 0.20$). We then tested whether all null alleles in both populations correspond to a single shared allele. Distributions of null allele sizes, within and between populations, were characterized for datasets harboring null alleles. This allowed us to estimate the within-population percentages of allele sizes associated with null gene copies for

the focal and the non focal populations and the percentage of allele sizes associated with null gene copies that are shared by both populations.

In the remaining tests, we simulated datasets of 10 and 100 loci. Researchers typically counter the large variances of differentiation estimators by examining between five and twenty loci. Ten loci thus mimic a typical empirical dataset. However, larger numbers of loci (e.g., several hundred) are required for reliable estimates of between-population parameters, such as migration rates (Whitlock and McCauley 1999) or times of population splitting events (Zhivotovsky and Feldman 1995). We assessed the effect of null alleles on population differentiation estimation by evaluating the Weir's (1996) unbiased estimator of $F_{ST}$, the genetic distance of Cavalli-Sforza and Edwards (1967) ($D_C$), and Nei's (1978) standard genetic distance ($D_S$). We compared the differentiation estimators for *VA* and *NA* datasets that correspond to the same set of parameters. We then estimated null allele frequencies averaged over the two populations, using the three methods of Dempster, Laird, and Rubin (1977; DEMPSTER method; estimate $\hat{r}_D$), Chakraborty et al. (1992; CHAKRABORTY method; estimate $\hat{r}_C$) and Brookfield (1996; BROOKFIELD method; estimate $\hat{r}_B$). Details about the null allele frequency estimates are provided as *Supplementary Material*. We evaluated the methods according to (i) their applicability, expressed as the percentage of times an estimate was successfully produced and (ii) a comparison of the means of estimated and simulated frequencies of null alleles averaged over the two populations.

Finally, we assessed the performance of available (*INA*) and new (*ENA*; for $F_{ST}$ only) methods for estimating population differentiation with datasets that included null alleles. The efficiency of correction for estimates of $F_{ST}$ was evaluated with respect to Weir's (1996) $F_{ST}$ values calculated with *VA* datasets or Li's (1976) equilibrium value $F_{ST} = \dfrac{1}{1+2N_e\left(2\mu_R + 2\dfrac{n_d}{n_d-1}m\right)}$, with the number of demes $n_d = 2$. As the two comparisons

gave similar results (details not shown), only the comparison with Weir's (1996) $F_{ST}$ values calculated with *VA* datasets is shown. As the relationship $D_S = 2\mu_R t$ (Nei 1972) does not hold under a GSM (Takezaki and Nei 1996), it was not considered in our comparisons. The performances of *INA* and *ENA* were evaluated by (i) comparing the distributions of each estimator of $F_{ST}$, $D_S$ and $D_C$ calculated from *CNA* datasets, according to *INA* and *ENA* for $F_{ST}$ and *INA* for $D_S$ and $D_C$, with those calculated from *VA* datasets and, (ii) calculating a success index for the corrections. This index corresponds to the percentage of times the differentiation estimate obtained with the *VA* dataset was closer to the differentiation estimate obtained with the *CNA* dataset, by *INA* or *ENA*, than to the differentiation estimate obtained with the *NA* dataset. For instance, for $F_{ST}$, we calculated the percentage of times $\left|\hat{F}_{ST[CNA]} - \hat{F}_{ST[VA]}\right| < \left|\hat{F}_{ST[NA]} - \hat{F}_{ST[VA]}\right|$.

Application to empirical molecular data

In some studies, the inference that null alleles are present leads to the design of new primers for PCR amplification of DNA from all individuals originally identified as homozygous or null (review in Dakin and Avise 2004). Although the two datasets obtained in this way are the empirical equivalents of our simulated *NA* and *VA* datasets, redesigning new primers does not guarantee that all null alleles are recovered (Ishibashi et al. 1996). To illustrate our simulation results with empirical molecular data, we re-analyzed two such published microsatellite datasets: a single locus from three Kenyan populations of the mosquito *Anopheles gambiae* (Lehmann, Hawley, and Collins 1996), and a single locus from three brown bear (*Ursus americanus*) populations sampled in Canadian National Parks (Paetkau and Strobeck 1995). These datasets represent different taxa, microsatellite loci, null allele frequencies, gene diversities, and levels of population differentiation. We first checked the recovery of HW equilibrium for each population using the genotype datasets obtained

with new primers (Fisher's exact tests as implemented in GENEPOP; Raymond and Rousset 1995). For each dataset, we then calculated an empirical null allele frequency as the frequency of gene copies amplified only with the new primers. We compared this empirical estimation with estimates of null allele frequency calculated from the original dataset, applying the three previously described methods. We compared global $F_{ST}$ and mean genetic distance statistics calculated from the original dataset, the new dataset, and the original dataset corrected for the presence of null alleles.

**Results**

Null allele prevalence and distribution

We first tested the prediction that genetic diversity in binding sites *B* (determined by $N_e \mu_B$ values) substantially affects null allele prevalence in the focal population (Figure 2, dotted line). For values of $N_e \mu_B$ below 0.001, the prevalence of null alleles was low for most loci ($r < 0.05$). For values of $N_e \mu_B$ greater than 0.1, the incidence of null alleles was high, with most loci having a high frequency of null alleles ($r \geq 0.20$ for 71% of loci). For intermediate values of $N_e \mu_B$, a substantial proportion of loci had a high frequency of null alleles ($r \geq 0.20$), and a moderate proportion of loci had an intermediate null allele frequency ($0.05 \leq r < 0.20$ for less than 19% of loci).

We then investigated how genetic differentiation from the focal population might favor null allele prevalence in the non-focal population. Gene flow had a low to moderate impact on null allele prevalence (Figure 2a). The focal and non-focal populations behaved similarly under high gene flow conditions ($N_e m = 10$). However, for low values of gene flow ($N_e m = 0.1$), the non-focal population was more strongly affected by null alleles. In the population split model, in which there was assumed to be no gene flow (Figure 2b), both populations had very similar distributions of loci harboring null alleles at various frequencies

for short to moderate splitting times ($t < 1,000$ generations). For longer times, the non-focal population was much more strongly affected by null alleles, even for low $N_e \mu_B$ values.

Finally, we investigated whether all null alleles in all populations correspond to a single shared allele size. Figure 3 shows the distribution of null alleles according to allele sizes, within and between populations. For both population models, a large number of allele sizes harbored null gene copies whatever the value of $N_e \mu_B$ (Figure 3a). In the migration model, the focal and non-focal populations behaved similarly for moderate to high levels of gene flow, with more than 34% of allele sizes harboring null gene copies. For low values of gene flow (i.e., $N_e m$ = 0.1), the non-focal population displayed a slightly higher number of allele sizes with null gene copies (results not shown). In the population split model, the non-focal population displayed a much larger number of allele sizes with null gene copies ($\geq 60\%$ for $t = 10,000$) than the focal population. This result held for a large range of splitting times (results not shown). In the migration model, less than half of the allele sizes harboring null gene copies were shared between the two populations for almost all combinations of parameter values tested (Figure 3b). The proportion of shared null allele sizes decreased with lower gene flow and $N_e \mu_B$. In the population split model, for all splitting times tested, populations shared very few allele sizes harboring null gene copies (less than 20% in most cases).

Effect of null alleles on the estimation of population differentiation

We tested the prediction that the presence of null alleles causes bias in differentiation estimators (Figure 4). The presence of null alleles led to overestimation of both $F_{ST}$ and genetic distance. In the migration model, bias in $F_{ST}$ was moderate for intermediate null allele frequencies or high levels of gene flow. Larger bias was observed for high null allele frequencies and low levels of gene flow, with the $F_{ST}$ distributions based on *VA* and *NA*

datasets becoming almost non-overlapping. In the population split model, the effect on genetic distances remained moderate, even for large null allele frequencies and large splitting times. $D_C$ was found to be slightly less affected by null alleles than $D_S$. $D_S$ could not be calculated for a diverse range of $N_e \mu_B$ and $t$ values (results not shown). These failures to calculate $D_S$ corresponded to paired populations that did not share at least one allele state. This situation is likely for large splitting times, even in the absence of null alleles. However, in the presence of null alleles, the probability of sharing no allele increases (results not shown). Increasing the number of loci reduced the variance of estimation for both $F_{ST}$ and genetic distances but did not change the null allele bias.

Estimation of null allele frequency

The performance of the methods for estimating null allele frequency under the migration model and for genotype datasets of 10 loci are presented in Figure 5. The results obtained for the population split model and for genotype datasets of 100 loci were similar and are therefore not shown. The CHAKRABORTY method generated negative estimates of null allele frequency (Figure 5a) when the simulated null allele frequency was close to zero for at least one of the two populations and when the number of visible genotypes for one population was too small for correct estimation of the observed heterozygozity. The CHAKRABORTY method was also not applicable for monomorphic populations. The CHAKRABORTY method gave a small positive bias and a large variance, especially for large values of null allele frequency (Figure 5b). This may simply result from sample size being reduced in this case because estimation with the CHAKRABORTY method is carried out for individuals with at least one visible band. Other methods had an applicability of 1 for all sets of parameter values tested. The BROOKFIELD method displayed a slight positive bias and its variance was low. The DEMPSTER method provided unbiased and low variance estimates of null allele frequencies.

Results were similar for a wide range of $N_em$ and $t$ values, and number of loci (results not shown). We therefore conclude that the DEMPSTER method was the best method of the three for estimating null allele frequencies.

Correction methods for estimating population differentiation

Figure 4 shows the differentiation estimates obtained from *CNA* datasets including (*INA*) or excluding (*ENA*) the null allele size for different categories of null allele frequency and numbers of loci. The *INA* correction continued to generate biased values of $F_{ST}$. This procedure partially educed the bias induced by null alleles in the presence of high levels of gene flow, generating values of $F_{ST}$ estimates smaller than those obtained with uncorrected datasets. However, this procedure increased the bias induced by null alleles in the presence of low levels of gene flow, with $F_{ST}$ estimates reaching values larger than those obtained from uncorrected datasets. In contrast, the newly proposed *ENA* method almost entirely resolved the bias induced by null alleles, regardless of null allele frequency, the level of gene flow, and the number of loci. Variance estimates for the *ENA* method were only slightly larger than those with *VA* datasets. These results were confirmed by success index values, which were larger than 67% for 10 loci and 95% for 100 loci (Figure 4).

*INA* decreased the bias in genetic distance estimation, almost eliminating it for moderate null allele frequencies. However, *INA* gave a negative bias for high null allele frequencies. These findings applied to both $D_S$ and $D_C$, but the bias was substantially less pronounced for $D_C$ than for $D_S$. For ten loci, *INA* only marginally improved genetic distance estimation, as confirmed by success index values (Figure 4a). Increasing the number of loci to 100 increased the success index values for *INA*, in spite of similar biases (e.g., between 68% and 96% for $D_C$; Figure 4b). This probably results from a much smaller variance of distance estimation for large number of loci. Thus, there appears to be a gain in using $D_C$ corrected by conventional methods, at least for datasets with a large number of loci.

Application to empirical molecular datasets

$H_O$ and $H_E$ values and tests of HW disequilibrium showed that null alleles were largely eliminated by the design of new primers for both *A. gambiae* and *U. americanus* (Table 1a). However, the heterozygote deficit remained significant for *A. gambiae*. As some null genotypes were still observed and Lehmann et al. (1997) excluded the Wahlund effect as an explanation of HW deviations in the genotype dataset obtained with the original primer set, the smaller, but still significant, HW deviation in the dataset obtained with the new primers may reflect the presence of non-recovered null alleles. Population estimates of null allele frequency $\hat{r}$ were generally close to the empirical values, estimated as the frequency of gene copies amplified only with the new primers. However, all $\hat{r}$ values were larger than the empirical $r$ values for *A. gambiae* populations, probably due to the incomplete recovery of null alleles in this species. Moreover, $\hat{r}_C$ values also appeared to be overestimated in *U. americanus*, especially for the population from Fundy, probably due to its small sample size.

The conclusions drawn from the population differentiation tests were the same for all three datasets (original primer dataset, new primer dataset, and corrected original primer dataset): no significant differentiation in *A. gambiae* populations and significant differentiation in *U. americanus* populations (results not shown). In agreement with our simulation results, $F_{ST}$ and genetic distances were considerably larger in the original dataset harboring null alleles than in the dataset obtained with the new primers, at least when genetic differentiation was significant (i.e., in *U. americanus*; Table 1b). The corrected dataset gave lower $D_S$ and $D_C$ values than the new primer dataset, consistent with simulation results. However, the $F_{ST}$ value obtained for *U. americanus* with the new primer dataset was more similar to that calculated from the original dataset than to that calculated from the corrected dataset. This may be due to the large variance observed in our simulations for single locus $F_{ST}$ estimation, regardless of the dataset considered (results not shown).

## Discussion

### Null allele prevalence

Our simulations showed that null alleles were likely to be encountered in populations with high levels of diversity in flanking sequences, particularly for $N_e \mu_B \geq 0.001$. Assuming a frequency of point mutations at a specific base pair of $10^{-9}$ (Li, Luo, and Wu 1985), the mutation rate in key regions of the binding sites for microsatellite primers (i.e., the 10 pb binding to the 3' end of each 20 pb-long primer), $\mu_B$, is expected to be about $2 \times 10^{-8}$. Hence, null alleles are likely to be found only in populations with large effective sizes (i.e., $N_e \geq 50,000$, and even larger population sizes if some mutations in the 10 pb binding sites do not preclude PCR amplification in spite of the primer mismatch to the DNA template). The prevalence of null alleles varies considerably between studies, but microsatellite null alleles have been found in a wide range of taxa, including species for which $N_e$ is not necessarily large (Dakin and Avise 2004). High mutation rates in the flanking sequences of microsatellite loci would be required to reconcile such empirical results with our simulations. In agreement with this, several molecular studies suggest that microsatellite flanking regions may be more unstable than is generally thought (Angers and Bernatchez 1997; Grimaldi and Crouau-Roy 1997; Meglecz et al. 2004). A simpler non exclusive explanation for the frequent presence of null alleles in most real datasets is the high level of differentiation that may exist between the focal population and the genotyped populations. In agreement with molecular studies (Li et al. 2003), our simulations showed that the non-focal population was more strongly affected by null alleles than the focal population, even for low $N_e \mu_B$ values.

### Effect of null alleles on the estimation of population differentiation

Simulated and empirical datasets showed that the presence of null alleles led to the overestimation of both $F_{ST}$ and genetic distance in cases of significant population differentiation. $F_{ST}$ estimates were unbiased in the absence of population structure but were considerably affected in the presence of low levels of gene flow (i.e., strongly differentiated populations). The presence of null alleles may be particularly problematic in studies comparing different sets of populations with different frequencies of null alleles and/or patterns of gene flow, especially when one or several population sets are characterized by low levels of gene flow. Cavalli-Sforza and Edwards' (1967) distance ($D_C$) performed better than Nei's (1978) standard distance ($D_S$): $D_C$ was less affected by null alleles and the bias remained similar for a large range of splitting times. This feature is important because genetic distances based on microsatellites are usually calculated for the construction of dendrograms of related taxa. If all pairwise $D_S$ distances are similarly biased then the tree topology should be roughly unchanged.

Correction methods for estimating population differentiation

Although the frequency of null alleles can be estimated precisely by the DEMPSTER method, the conventional correction based upon this estimate of null allele frequency did not perform well. Bias in $F_{ST}$ is larger after correction for null alleles in the presence of low levels of gene flow. Genetic distances calculated from corrected datasets were underestimated when null allele frequencies were high. However, the absolute bias on Cavalli-Sforza and Edwards' (1967) distance was lower than for uncorrected datasets. Our simulations demonstrated that null alleles often corresponded to multiple allele sizes, some of which were similar to those of visible alleles. This is due to the mutational model of the repeat region of the microsatellite, in which the loss or gain of a variable number of repeat units generates alleles identical in state but not in descent (i.e., allele size homoplasy; Estoup, Jarne, and Cornuet 2002). This issue

was more pronounced in higher levels of population differentiation where population differences in allele sizes of null gene copies were larger. The conventional assumption of a single null allele size common to all studied populations, rather than the actual allele sizes, amounts to considering these alleles as slowly evolving, and so decreases the apparent overall mutation rate of the locus. As $F_{ST}$ increases with decreasing $N_e\mu_R$ (Slatkin 1995), we would expect $F_{ST}$ values calculated with the *INA* procedure to be overestimated with respect to $F_{ST}$ values calculated from *VA* datasets (particularly in low gene flow conditions). Conversely, as genetic distance decreases with decreasing $\mu_R t$ (Nei 1972), the genetic distances values calculated with the *INA* procedure should be lower than those calculated from *VA* datasets. The assumption of arbitrarily choosing a single allele size common to all null alleles can be relaxed, at least when estimating $F_{ST}$, by restricting $F_{ST}$ calculation from corrected datasets to visible allele sizes. $F_{ST}$ calculation with the *ENA* procedure was unbiased and resulted in a variance only slightly larger than that for datasets without null alleles.

**Acknowledgments**

We would like to thank T. Lehmann and D. Paetkau for providing us with the datasets on which their publications were based. We thank S. Baird, D. Bourguet, C. Brouat, J.M. Cornuet, K. Kim, Y. Michalakis, G. Roderick, T. Sappington, and two anonymous reviewers for constructive comments on an earlier version of the manuscript. This work was partly funded by the scientific SPE department of INRA. M.P.C. was supported by a grant from the CNRS.

**Literature Cited**

Angers, B., and L. Bernatchez. 1997. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol. Biol. Evol. **14**:230-238.

Astanei, I., E. Gosling, J. Wilson, and E. Powell. 2005. Genetic variability and phylogeography of the invasive zebra mussel, *Dreissena polymorpha* (Pallas). Mol. Ecol. **14**:1655-1666.

Brookfield, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Mol. Ecol. **5**:453-455.

Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards, and J. C. Mulley. 1993. Incidence and origin of 'null' alleles in the (AC)*n* microsatellite markers. Am. J. Hum. Genet. **52**:922-927.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Genet. **19**:233-257.

Chakraborty, R., M. De Andrade, S. P. Daiger, and B. Budowle. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann. Hum. Genet. **56**:45-57.

Chakraborty, R., M. Kimmel, D. N. Stivers, L. J. Davison, and R. Deka. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA **94**:1041-1046.

Chapuis, M.-P., A. Loiseau, Y. Michalakis, M. Lecoq, and A. Estoup. 2005. Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*. Mol. Ecol. Notes **5**:554-557.

Dakin, E. E., and J. C. Avise. 2004. Microsatellite null alleles in parentage analysis. Heredity **93**:504-509.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J Roy. Stat. Soc. B **39**:1-38.

Estoup, A., and B. Angers. 1998. Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. Pp. 55-86 *in* G. Carvalho, ed. Advances in Molecular Ecology. NATO ASI series. IOS Press, Amsterdam.

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consquences for population genetics analysis. Mol. Ecol. **11**:1591-1604.

Estoup, A., I. J. Wilson, C. Sullivan, J. M. Cornuet, and C. Moritz. 2001. Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. Genetics **159**:1671-1687.

Gagneux, P., C. Boesch, and D. S. Woodruff. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. Mol. Ecol. **6**:861-868.

Grimaldi, M. C., and B. Crouau-Roy. 1997. Microsatellite allelic homoplasy due to variable flanking sequences. J Mol. Evol. **44**:336-340.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. Pp. 1-44 *in* D. Futuyama and J. Antonovics, eds. Oxford Surveys in Evolutionary Biology. Oxford University Press, Oxford.

Ishibashi, Y., T. Saitoh, S. Abe, and M. C. Yoshida. 1996. Null microsatellite alleles due to nucleotide sequence variation in the grey-sided vole *Cleithrionomy rufocanus*. Mol. Ecol. **5**:589-590.

Jin, L., C. Macaubas, J. Hallmayer, A. Kimura, and E. Mignot. 1996. Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. Proc. Natl. Acad. Sci. USA **93**:15285-15288.

Kimura, M., and J. F. Crow. 1964. The number of alleles that can be maintained in a finite population. Genetics **49**:725-738.

Leblois, R., A. Estoup, and F. Rousset. 2003. Influence of mutational and sampling factors on the estimation of demographic parameters in a "Continuous" population under isolation by distance. Mol. Biol. Evol. **20**:491-502.

Lehmann, T., N. J. Besanky, W. A. Hawley, T. G. Fahey, L. Kamau, and F. H. Collins. 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. Mol. Ecol. **6**:243-253.

Lehmann, T., W. A. Hawley, and F. H. Collins. 1996. An evaluation of evolutionary constraints on microsatellite loci using null alleles. Genetics **144**:1155-1163.

Li, G., S. Hubert, K. Bucklin, V. Ribes, and D. Hedgecock. 2003. Characterization of 79 microsatellite DNA markers in the Pacific oyster *Crassostrea gigas*. Mol. Ecol. Notes **3**:228-232.

Li, W.-H., C.-C. Luo, and C.-I. Wu. 1985. Evolution of DNA sequences. Pp. 1-94 *in* R.J. Macintryre, ed. Molecular Evolutionary Genetics. Plenum Press, New-York.

Meglecz, E., F. Petenian, E. Danchin, A. Coeur d'Acier, J. Y. Rasplus, and E. Faure. 2004. High similarity between flanking regions of different microsatellites detected within each of two species of lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. Mol. Ecol. **13**:1693-1700.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**:283-291.

———. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89**:583-590.

Paetkau, A., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol. Ecol. **13**:55-65.

Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. Mol. Ecol. **4**:519-520.

Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead, and C. Strobeck. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics **147**:1943-1957.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Heredity **86**:248-249.

Roques, S., P. Duchesne, and L. Bernatchez. 1999. Potential of microsatellites for individual assignment: the North Atlantic redfish (genus *Sebastes*) species complex as a case study. Mol. Ecol. **8**:1703-1717.

Shinde, D., Y. L. Lai, F. Z. Sun, and N. Arnheim. 2003. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites. Nucl. Ac. Res. **31**:974-980.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**:457-462.

Takezaki, N., and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics **144**:389-399.

Weir, B. S. 1996. Genetic Data Analysis II. Sinauer Associates, Sunderland, Mass.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{ST}$ not equal 1/(4$Nm$+1). Heredity **82**:117-125.

Zhivotovsky, L. A., and M. W. Feldman. 1995. Microsatellite variability and genetic distances. Proc. Natl. Acad. Sci. USA **92**:11549-11552.

Zhivotovsky, L. A., M. W. Feldman, and S. A. Grishechkin. 1997. Biased mutations and microsatellite variation. Mol. Biol. Evol. **14**:926-933.

**Table**

**Table 1. Null alleles in empirical molecular data**

**(a) Dataset details and estimation of null allele frequency**

| Sample site | *n* | Original primer set | | | | New primer set | | | | Null allele frequencies | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $n_0$ | $H_O$ | $H_E$ | *HW test* | $n_0$ | $H_O$ | $H_E$ | *HW test* | *r* | $\hat{r}_C$ | $\hat{r}_B$ | $\hat{r}_D$ |
| *A. gambiae*[(a)] | | | | | | | | | | | | | |
| Village 3 | 39 | 7 | 0.415 | 0.860 | * | 4 | 0.705 | 0.870 | * | 0.174 | 0.344 | 0.427 | 0.367 |
| Village 7 | 54 | 3 | 0.352 | 0.854 | * | 0 | 0.737 | 0.867 | * | 0.184 | 0.412 | 0.344 | 0.316 |
| Village 15 | 70 | 7 | 0.378 | 0.848 | * | 3 | 0.731 | 0.860 | * | 0.224 | 0.380 | 0.373 | 0.331 |
| *U. americanus*[(b)] | | | | | | | | | | | | | |
| Fundy | 11 | 3 | 0.000 | 0.600 | * | 0 | 0.636 | 0.502 | *ns* | 0.591 | 1.000 | 0.643 | 0.589 |
| La Mauricie | 31 | 2 | 0.379 | 0.878 | * | 0 | 0.774 | 0.856 | *ns* | 0.242 | 0.389 | 0.351 | 0.322 |
| Terra Nova | 26 | 1 | 0.080 | 0.520 | * | 0 | 0.385 | 0.529 | *ns* | 0.192 | 0.729 | 0.351 | 0.336 |

NOTE—Sample size in diploid individuals (*n*), number of null genotypes ( $n_0$ ), observed ($H_O$) and expected ($H_E$) heterozygosities for original and new primer sets. Null allele frequencies in original datasets were calculated as described by CHAKRABORTY: $\hat{r}_C$ , BROOKFIELD: $\hat{r}_B$ and DEMPSTER: $\hat{r}_D$ (see Appendix 1). *r* is the 'real' estimate of null allele frequency calculated as the frequency of genes amplified only with new primers. Datasets were originally published in (a) Lehmann, Hawley, and Collins (1996) and (b) Paetkau and Strobeck (1995). *HW test*: Hardy-Weinberg test for heterozygote deficiency (Raymond and Rousset 1995), *: significant departure at $\alpha = 0.05$; *ns*: not significant.

**(b) Estimation of genetic differentiation**

| Differentiation estimator | | Population set | Original primer dataset | New primer dataset | Original primer dataset corrected by | |
|---|---|---|---|---|---|---|
| | | | | | *INA* | *ENA* |
| Global | $F_{ST}$ | *A. gambiae* | -0.011 | -0.005 | -0.005 | -0.005 |
| | | *U. americanus* | 0.177 | 0.150 | 0.078 | 0.092 |
| Mean | $D_S$ | *A. gambiae* | -0.036 | -0.026 | -0.019 | *na* |
| | | *U. americanus* | 0.727 | 0.354 | 0.234 | *na* |
| Mean | $D_C$ | *A. gambiae* | 0.122 | 0.135 | 0.110 | *na* |
| | | *U. americanus* | 0.566 | 0.498 | 0.445 | *na* |

NOTE—Original datasets were corrected using $\hat{r}_D$ estimation. $D_S$: Nei's (1978) standard distance; $D_C$: Cavalli-Sforza and Edwards' (1967) distance; *INA*: calculation of the differentiation measures ($F_{ST}$ and genetic distance) from the dataset corrected for null alleles when the null allele size is included; *ENA*: calculation of the $F_{ST}$ from the dataset corrected for null alleles when the null allele size is excluded; *na*: not applicable.

**Figure legends**

**FIG. 1.—Synopsis of the simulation method.**

NOTE—A single iteration is presented. In the coalescent tree, the allele state in the binding sites of the two microsatellite primers of the "grey" gene copies leads to null alleles. Estimation of genetic differentiation is illustrated by estimation of $F_{ST}$. $\hat{\sigma}_P^2$, $\hat{\sigma}_I^2$, and $\hat{\sigma}_G^2$ are the estimated components of variance for populations, individuals within populations, and genes within individuals, respectively. GSM: generalized stepwise mutational model (Zhivotovsky et al. 1997 and Estoup et al. 2002); IAM: infinite allele model (Kimura and Crow 1964); *R*: repeat region; *B*: primer-binding sites.

**FIG. 2.—Prevalence of null alleles.**

NOTE—Frequencies of simulated loci with a null allele frequency $r < 0.05$ (light gray), $0.05 \leq r < 0.20$ (dark gray) and $r \geq 0.20$ (black) as a function of the parameter $N_e \mu_B$ (*x* axis). Dotted lines represent the focal population and solid lines represent the non-focal population. Different levels of gene flow and splitting time are tested for a migration model (a) and a model of population divergence (b).

**FIG. 3.—Allele sizes harboring null gene copies.**

NOTE—Distribution of null allele sizes within (a) and between (b) populations, calculated only for datasets including null alleles and presented along the *y* axis as a function of the parameter $N_e \mu_B$ (*x* axis). *P:* proportion within population of allele sizes harboring null gene copies; *Ps:* proportion of allele sizes harboring null gene copies that are shared by both populations. Mean estimates (line), 50 (points), 10 and 90 (bars) % quantile values are represented. (a) Both the focal (dotted line) and the non-focal (solid line) populations are presented. The parameter $N_e m$ is fixed at 1 for the migration model. The parameter *t* is fixed at 10,000 for the model of population divergence. (b) For the migration model, the tested

values of gene flow are $N_em$ = 0.1 (light gray), $N_em$ = 1 (dark gray) and $N_em$ = 10 (black). For the model of population divergence, the tested values of splitting time are $t$ = 100,000 (light gray), $t$ = 10,000 (dark gray) and $t$ = 1,000 (black).

**FIG. 4.—Effects of null alleles on estimation of population differentiation and performance of correction methods for a genotyping effort of 10 loci (a) and 100 loci (b).**
NOTE—$F_{ST}$ and genetic distance estimates ($y$ axis) are presented as a function of gene flow ($N_em$) for $F_{ST}$ or splitting time ($t$) for genetic distances ($x$ axis). The differentiation estimates are based on *VA* datasets (black line), *NA* datasets (gray line) and *CNA* datasets including (*INA*, orange line) or excluding (*ENA*, blue line) the null allele size. Null allele frequency is estimated using the DEMPSTER method. Mean estimates (line), 50 (points), 10 and 90 (bars) % quantile values are represented. Numbers refer to success indices corresponding to the percentages of differentiation estimates based on the *VA* datasets that are closer to the differentiation estimates based on the *CNA* datasets than to the differentiation estimates based on the *NA* datasets. The *CNA* dataset estimate was generated following the *INA* (orange) or *ENA* (blue) correction method. All estimates were calculated for two classes of mean null allele frequency $\bar{r}$ : $0.05 \leq \bar{r} < 0.20$ and $\bar{r} \geq 0.20$. $D_S$: Nei's (1978) standard distance; $D_C$: Cavalli-Sforza and Edwards' (1967) distance.

**FIG. 5.—Performance of methods for estimating null allele frequency.**
NOTE—Applicability (a) and mean and quantile values (b) of null allele frequency estimates ($y$ axis) were plotted as a function of the simulated mean null allele frequency $\bar{r}$ ($x$ axis) grouped into classes of 0.1 units. The methods evaluated are those of CHAKRABORTY ($\hat{r}_C$ : □, black), BROOKFIELD ($\hat{r}_B$: ○, dark gray) and DEMPSTER ($\hat{r}_D$ : △, light gray), as described in *Supplementary Material*. Calculations were performed under the migration model and for genotype datasets of 10 loci. (a) The applicability is the percentage of times an estimate is

successfully produced. Different values of gene flow were tested: $N_em$ = 0.1 (solid line), $N_em$ = 1 (broken line) and $N_em$ = 10 (dotted line). For $\hat{r}_B$ and $\hat{r}_D$, the corresponding different lines are merged. (b) Mean estimates (lines), 50 (points), 10 and 90 (bars) % quantile values are presented. $N_em$ was fixed at 1.

**FIG. 1.**

Parameter values
$N_e \mu_R$, $N_e \mu_B$, $N_e m$ or $t$

↓

Coalescent tree of two samples of genes drawn from different populations under a given population model (migration or split)

↓

Attribution of allele states → GSM for *R* and IAM for *B*

↓

***VA* dataset** → generate genotypes for the repeat region assuming HW equilibrium (absence of null alleles)

↓

Primer sequence choice → draw a gene copy and a corresponding *B* state in the focal population (e.g. a black state)

↓

***NA* dataset** → repeat region allele states not associated with the drawn *B* state become null alleles

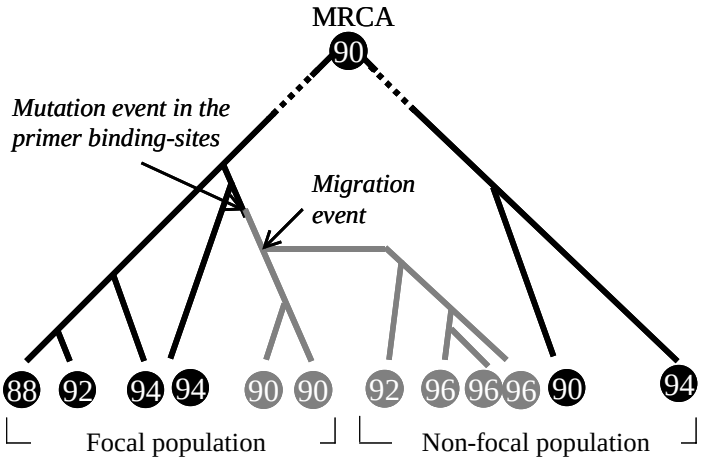
↓

***CNA* dataset** → estimate null allele frequency ($\hat{r}$), choose a single new allele state for null alleles (e.g., 99), and adjust homozygous *NA* genotypes using $\hat{r}$

↓

$F_{ST}$ estimation from datasets *VA*, *NA* and C*NA* (*INA* and *ENA* corrections)

## STEP 1. Simulation of allele states in the repeat region *R* and the primer-binding sites *B*




e.g., 94 5'-**TGTAAACGACGGCCAGTGAA**…$(GT)_n$…**AAAGAAGGAGGATTTAG**-5'

94 pb

## STEP 2. Design of the *NA*, *VA* and *CNA* datasets

| ***VA dataset*** | |
|---|---|
| Focal Population | Non-Focal Population |
| Ind1, 9092 | Ind4, 9096 |
| Ind2, 8894 | Ind5, 9696 |
| Ind3, 9094 | Ind6, 9294 |

| ***NA dataset*** | |
|---|---|
| Focal Population | Non-Focal Population |
| Ind1, 9292 | Ind4, 9090 |
| Ind2, 8894 | Ind5, 0000 |
| Ind3, 9494 | Ind6, 9494 |

| ***CNA dataset*** | |
|---|---|
| Focal Population | Non-Focal Population |
| Ind1, 9992 | Ind4, 9099 |
| Ind2, 8894 | Ind5, 9999 |
| Ind3, 9994 | Ind6, 9994 |

## STEP 3. Genetic differentiation estimation

$$\hat{F}_{ST}[\boldsymbol{VA}] = \frac{\sum_{i=88}^{i=96} \hat{\sigma}^2_{P(i)}}{\sum_{i=88}^{i=96} \left( \hat{\sigma}^2_{G(i)} + \hat{\sigma}^2_{I(i)} + \hat{\sigma}^2_{P(i)} \right)}, \hat{F}_{ST}[\boldsymbol{NA}] = \frac{\sum_{i=88}^{i=94} \hat{\sigma}^2_{P(i)}}{\sum_{i=88}^{i=94} \left( \hat{\sigma}^2_{G(i)} + \hat{\sigma}^2_{I(i)} + \hat{\sigma}^2_{P(i)} \right)}$$

$$\hat{F}_{ST}[\boldsymbol{CNA}]_{INA} = \frac{\sum_{i=88}^{i=99} \hat{\sigma}^2_{P(i)}}{\sum_{i=88}^{i=99} \left( \hat{\sigma}^2_{G(i)} + \hat{\sigma}^2_{I(i)} + \hat{\sigma}^2_{P(i)} \right)}, \hat{F}_{ST}[\boldsymbol{CNA}]_{ENA} = \frac{\sum_{i=88}^{i=94} \hat{\sigma}^2_{P(i)}}{\sum_{i=88}^{i=94} \left( \hat{\sigma}^2_{G(i)} + \hat{\sigma}^2_{I(i)} + \hat{\sigma}^2_{P(i)} \right)}$$

**FIG. 2.**

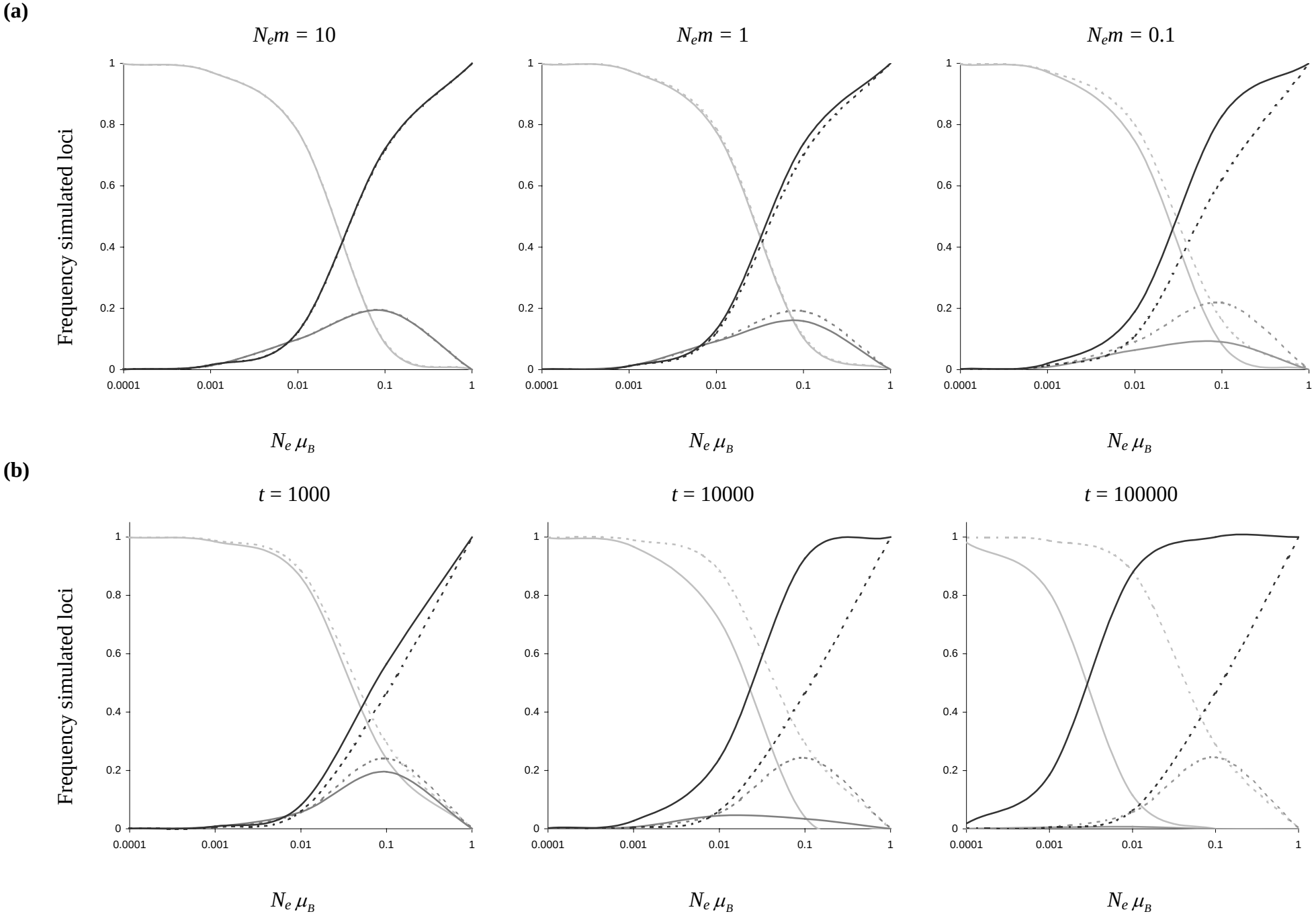
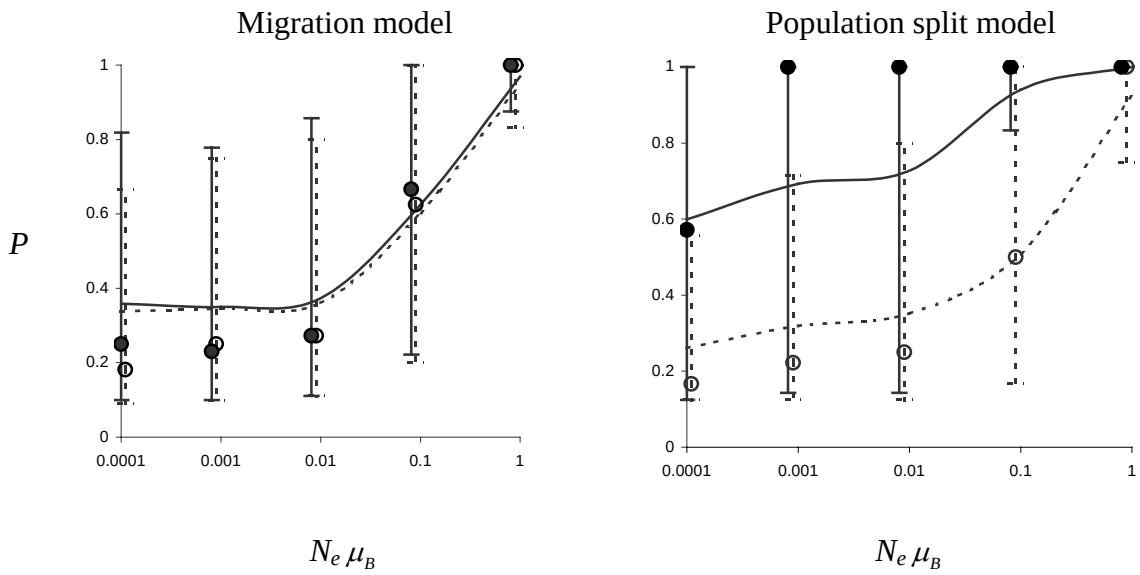

(a)
$N_e m$ = 10
$N_e m$ = 1
$N_e m$ = 0.1
Frequency simulated loci
0
0.2
0.4
0.6
0.8
1
0.0001
0.001
0.01
0.1
1
$N_e \mu_B$
(b)
$t$ = 1000
$t$ = 10000
$t$ = 100000
Frequency simulated loci
$N_e \mu_B$

**FIG. 3.**

**(a)**


Migration model
Population split model
P
1
0.8
0.6
0.4
0.2
0
0.0001
0.001
0.01
0.1
1
$N_e \mu_B$

**(b)**

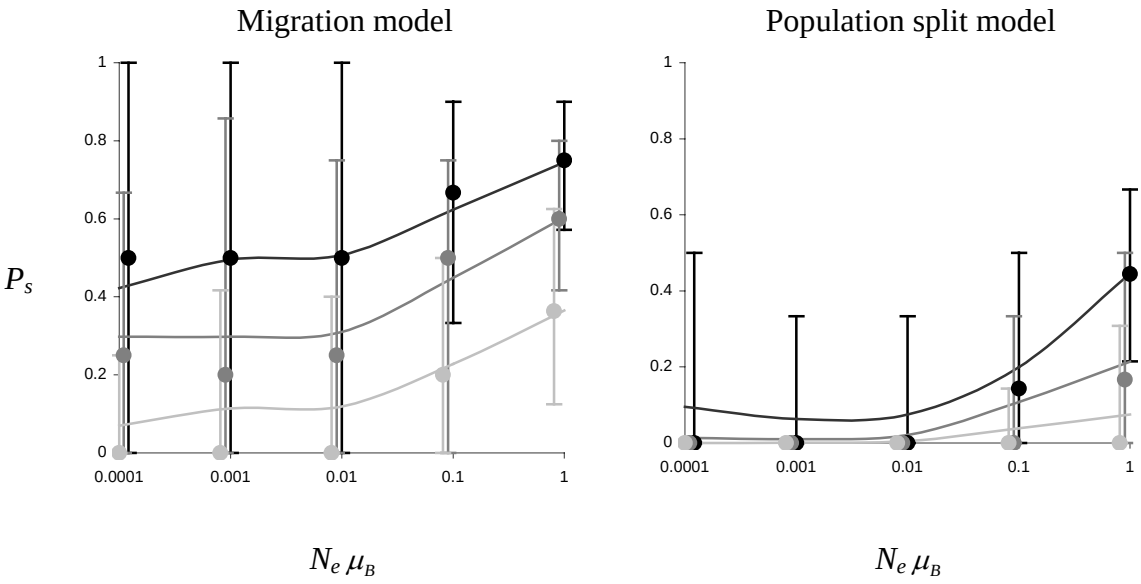
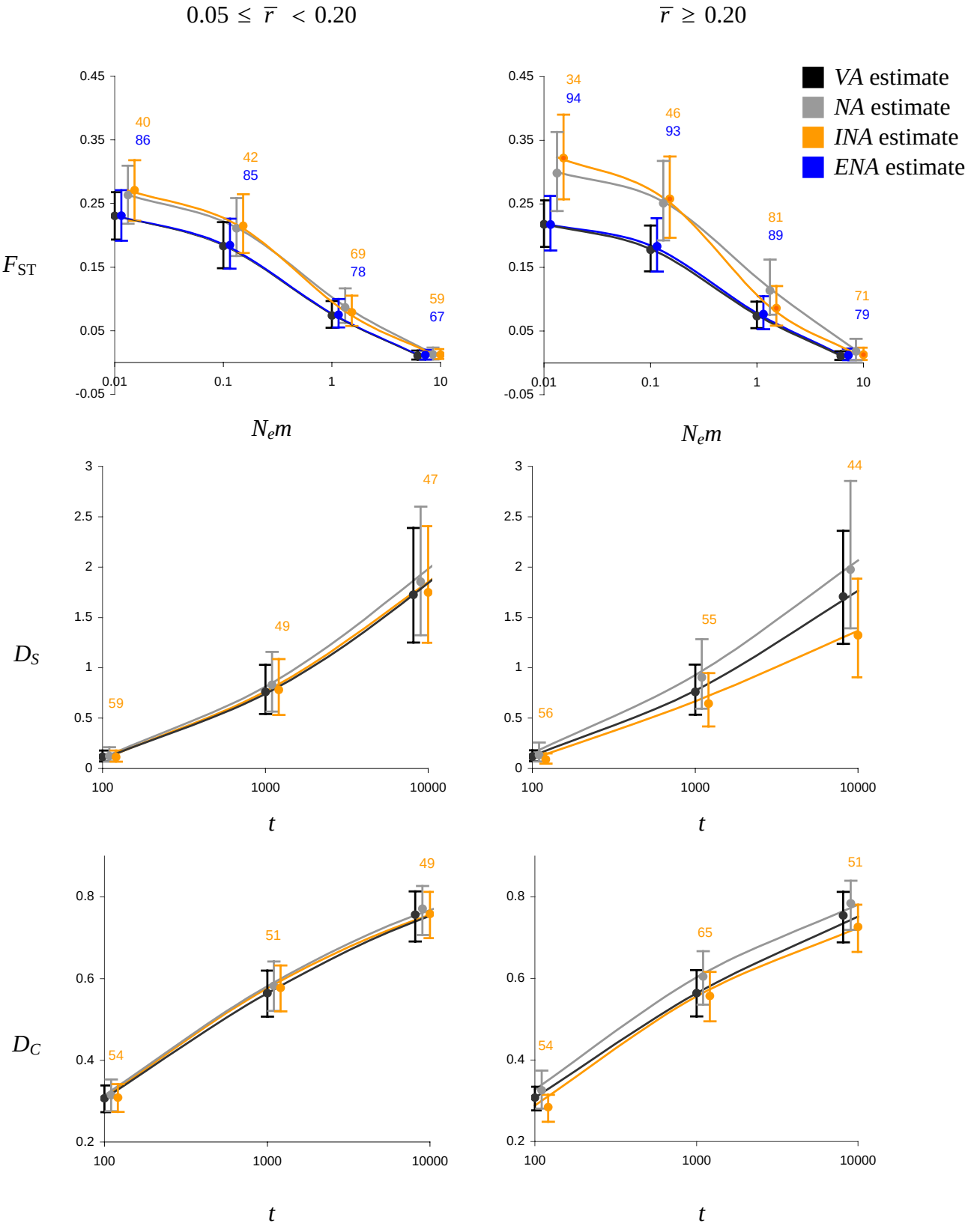

Migration model
Population split model
$P_s$
1
0.8
0.6
0.4
0.2
0
0.0001
0.001
0.01
0.1
1
$N_e \mu_B$

**FIG. 4.**

**(a)** 10 loci



**(b)** 100 loci

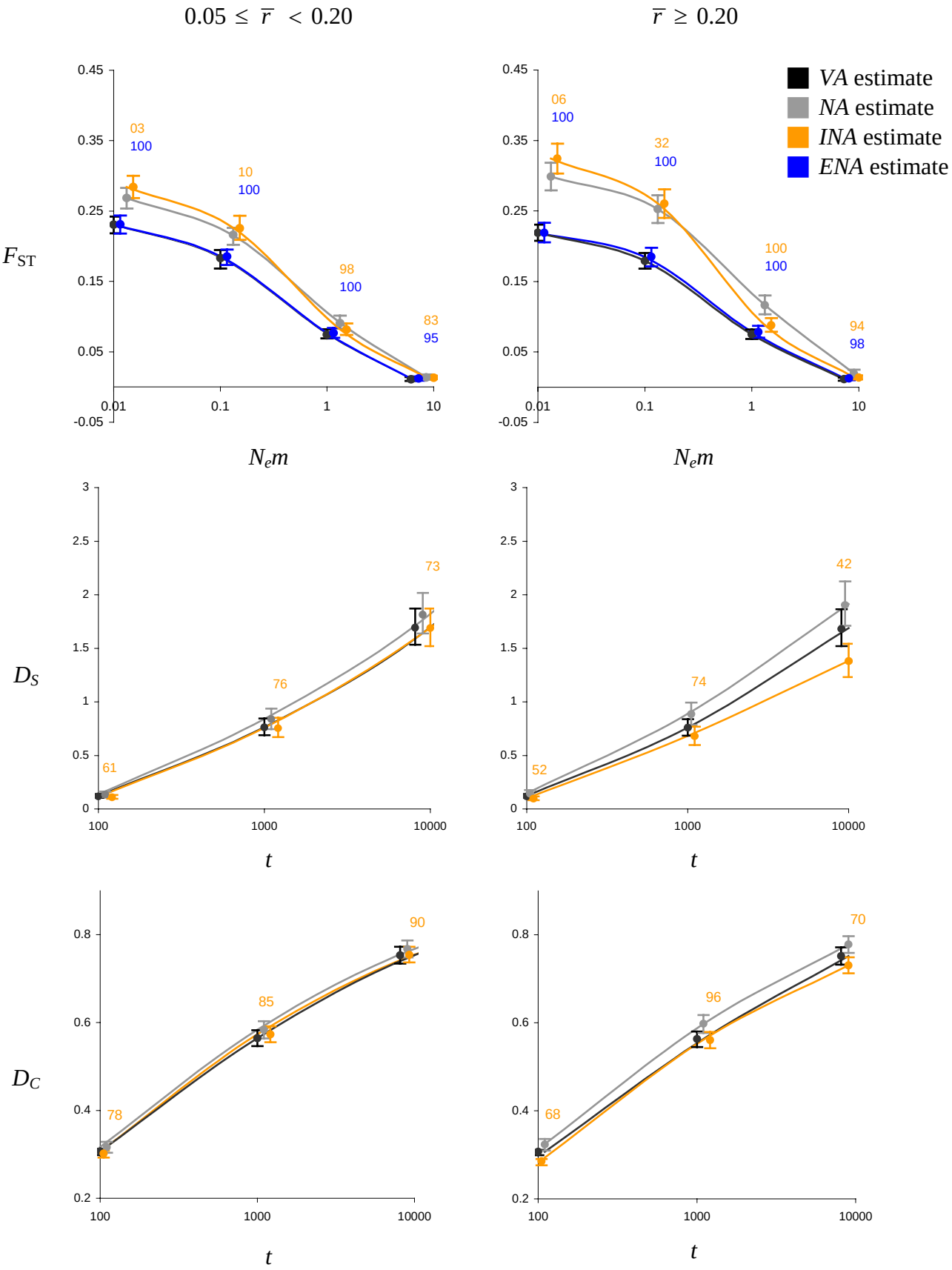
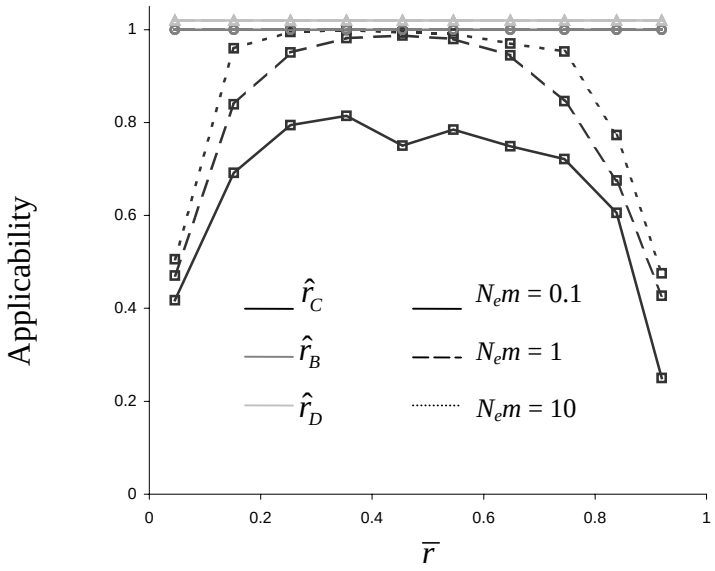
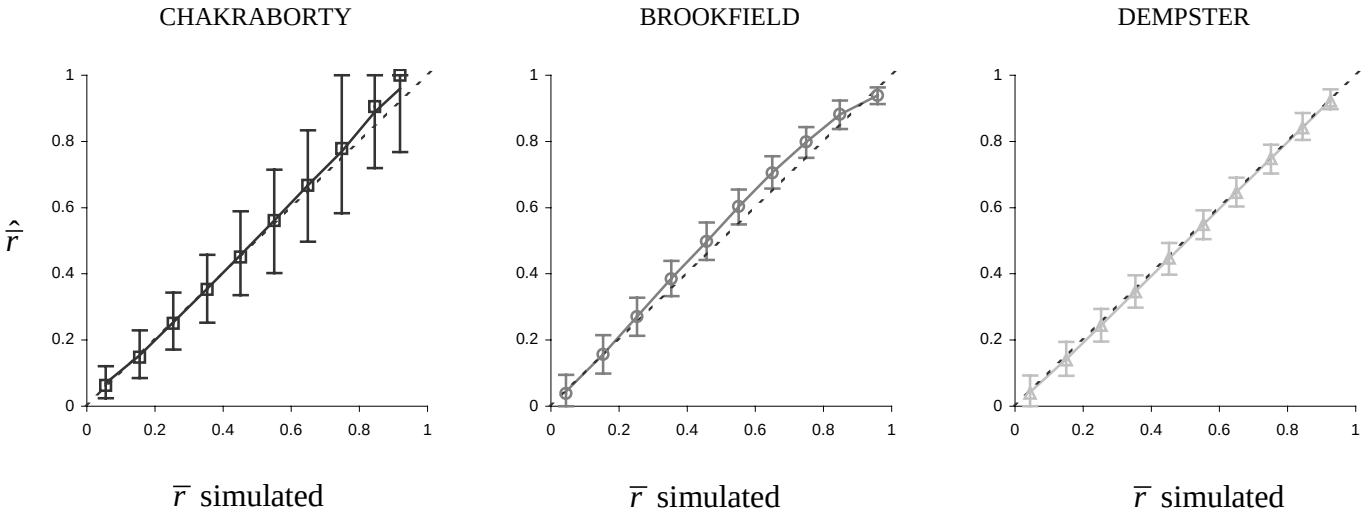

0.05 ≤ r̄ < 0.20
r̄ ≥ 0.20
VA estimate
NA estimate
INA estimate
ENA estimate
F_ST
N_em
D_S
D_C
t
03
100
10
100
98
100
83
95
06
100
32
100
100
100
94
98
61
76
73
52
74
42
78
85
90
68
96
70

**FIG. 5.**

**(a)**


Applicability
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
$\bar{r}$
$\hat{r}_C$
$\hat{r}_B$
$\hat{r}_D$
$N_e m = 0.1$
$N_e m = 1$
$N_e m = 10$

**(b)**


CHAKRABORTY
BROOKFIELD
DEMPSTER
$\hat{\bar{r}}$
$\bar{r}$ simulated
$\bar{r}$ simulated
$\bar{r}$ simulated

***Supplementary material***

**Methods for estimating null allele frequency**

Various methods for estimating null allele frequency from experimental genotype data have been developed. All the methods assume panmixia with all heterozygote deficiencies from HW proportions due to the presence of null alleles. The methods differ in the interpretation of individual samples without visible bands ($n_0$), which may be interpreted as individuals homozygous for a null allele (null homozygotes) or as PCR artifacts due to human error or degraded or low-quantity DNA. As our simulations were designed to address the issue of primer mismatch, we considered here only methods interpreting a lack of visible bands as corresponding to null homozygotes.

Let us consider a genotype dataset for one population of *n* individuals, including $n_0$ individuals with no bands (null genotypes). One visible allele $A_i$, of a total of *K* visible alleles, has an observed (apparent) frequency $p_i$, an actual frequency $p^*_i$, an observed number of homozygous individuals $n_{ii}$, and an observed number of heterozygous individuals $\sum_{j\neq i}^{K} n_{ij}$ (where *j* is a visible allele different from *i* and from null). $A_0$ is a null allele and *r* is the null allele frequency.

***- Estimation using the expectation-maximization (EM) algorithm (Dempster et al. 1977)***

In a genotype dataset with null alleles, individuals heterozygous for a null allele and a visible allele are scored as homozygous for the visible allele. Observed homozygous genotypes ($n_{ii}$) are therefore partitioned into true ($n^*_{ii}$) and false ($n^*_{i0}$) homozygous genotypes. The probability that the genotypes $A_iA_i$, $A_iA_j$ and $A_0A_0$ are observed in the dataset are hence $p^{*2}_i + 2p^*_i r$, $2p^*_i p^*_j$ and $r^2$, respectively. As a result, the likelihood of the genotype dataset is

$L = \left[\prod_{i=1}^{K}\left(p^{*}{}_{i}^{2} + 2p^{*}{}_{i}r\right)^{n_{ii}}\right]\left[\prod_{i=1}^{K}\left(2p^{*}{}_{i}p^{*}{}_{j}\right)^{\sum_{j\neq i}^{K} n_{ij}}\right]\left[\left(r^{2}\right)^{n_{0}}\right]$. As estimates of the visible allele frequencies $\hat{p}_i$ and null allele frequency $\hat{r}$ that maximize this likelihood cannot be found analytically, an EM approach is adopted.

The EM algorithm has commonly been used to find numerically the maximum likelihood estimates of allele frequencies in cases in which the observations are considered to be incomplete. The human *ABO* blood group system is a classic situation in which the procedure is used (Weir 1996). This system is similar to that for microsatellite null alleles because the recessive *O* allele is a null allele at the *ABO* blood protein locus. Consequently, the EM approach can easily be extended to estimation of the frequency of null alleles at a microsatellite locus (with *K* alleles instead of three in the *ABO* system).

The EM algorithm uses approximate frequency estimates $p'_i$ and $r'$ to calculate new estimates $p''_i$ and $r''$. The new estimates are then used as approximate estimates, and further new estimates are obtained. The EM algorithm converges on the maximum likelihood estimates provided that it is appropriately initialized ($r \neq 0$) and iterated using relevant equations (Dempster et al. 1977).

The initial frequency estimates were based on the work of Bernstein (1925; in Weir 1996):

$$\begin{cases} r' = \sqrt{\dfrac{n_0}{n}} \\ p'_i = 1 - \sqrt{\dfrac{\left(n_0 + \sum_{i=1}^{K} n_{ll} + \sum_{l\neq j}^{K} n_{lj}\right)}{n}} \end{cases} \text{ when } n_0 > 0$$

and

$$\begin{cases} r' = \sqrt{\dfrac{1}{n+1}} \\ p'_i = 1 - \sqrt{\dfrac{\left(1 + \sum_{i=1}^{K} n_{ll} + \sum_{l\neq j}^{K} n_{lj}\right)}{n+1}} \end{cases} \text{ when } n_0 = 0\text{, with } l \text{ a visible allele distinct from } i \text{ and null.}$$

The iterative equations are defined as follows. The true frequencies for each allele $A_i$ ($p^*_i$) and for the null allele ($r$), expressed as a function of genotype counts, are equal to: $p^*_i = \frac{1}{2n}\left[2n^*_{ii} + n^*_{i0} + \sum_{j\neq i}^{K} n_{ij}\right]$ and $r = \frac{1}{2n}\left[2n_0 + \sum_{i=1}^{K} n^*_{i0}\right]$. Now let us consider that the frequencies found in this way are the new estimates $p''_i$ and $r''$. Conversely, true and false genotype counts for each allele $A_i$ ($n^*_{ii}$ and $n^*_{i0}$) expressed as a function of true visible and null frequencies are equal to: $n^*_{ii} = np^{*2}_i$ and $n^*_{i0} = 2np^*_i r$. It is thought that the frequencies used in these relationships are the approximate estimates $p'_i$ and $r'$.

The iterative equations are:

$$p''_i = \frac{1}{2n}\left[2\frac{n^*_{ii}}{n_{ii}} \times n_{ii} + \frac{n^*_{i0}}{\times n_{ii}} \times n_{ii} + \sum_{j\neq i}^{K} n_{ij}\right] \text{ and } r'' = \frac{1}{2n}\left[2n_0 + \sum_{i=1}^{K} \frac{n^*_{i0}}{n_{ii}} \times n_{ii}\right].$$

Thus,

$$p''_i = \frac{1}{2n}\left[\frac{2p'^2_i + 2p'_i r'}{p'^2_i + 2p'_i r'} \times n_{ii} + \sum_{j\neq i}^{K} n_{ij}\right] \text{ and } r'' = \frac{1}{2n}\left[2n_0 + \sum_{i=1}^{K}\left(\frac{2p'_i r'}{p'^2_i + 2p'_i r'}\right) \times n_{ii}\right].$$

These equations can be slightly simplified:

$$p''_i = \frac{1}{2n}\left[\frac{2(p'_i + r')}{p'_i + 2r'} \times n_{ii} + \sum_{j\neq i}^{K} n_{ij}\right] \text{ and } r'' = \frac{1}{2n}\left[2n_0 + \sum_{i=1}^{K}\left(\frac{2r'}{p'_i + 2r'}\right) \times n_{ii}\right].$$

Iterations are continued until successive values are equal or almost so. For our simulation study, the EM algorithm was stopped when the quantities $(r'' - r')$ and $(p''_i - p'_i)$ for each allele $A_i$ were $< 10^{-5}$.

***- Estimation following the method of Chakraborty et al. (1992)***

In this method, individuals without bands are ignored, even though they may represent null homozygotes, and so the value of *r* is calculated for individuals with at least one band ($n - n_0$). Chakraborty et al. (1992) suggested that not counting absent alleles (of frequency *r*) distorted

the sum of visible allele frequencies by a factor 1 – $r$ and the sum of visible genotype frequencies by a factor 1 – $r^2$. They therefore expressed the observed heterozygote proportion $H_O$ and the expected heterozygosity $H_E$ as: $H_O = \frac{2\sum_{i\neq j}^{K} p^*_i p^*_j}{(1-r^2)}$ and $H_E = \frac{2\sum_{i\neq j}^{K} p^*_i p^*_j}{(1-r)^2}$. The null allele frequency is therefore estimated as $\hat{r}_C = \frac{H_E - H_O}{H_E + H_O}$.

***- Estimation following the method of Brookfield (1996)***

This method assumes that observed null genotypes correspond to individuals homozygous for a null allele and that the true frequency of the allele state $A_i$ is $p^*_i = p_i(1-r)$. A maximum likelihood approach is used to maximize the binomial distribution of genotype frequencies, summarized as follows: the expected frequency of heterozygotes $\frac{n^*_1}{n} = 2\sum_{i\neq j}^{K} p^*_i p^*_j = H_E(1-r)^2$, the expected frequency of null homozygotes $\frac{n^*_0}{n} = r^2$ and the expected frequency of homozygotes $\frac{n^*_2}{n} = 1 - \frac{n^*_1}{n} - \frac{n^*_0}{n}$. The maximum likelihood value of the null allele frequency is calculated as $\hat{r}_B = \frac{A + \sqrt{A^2 + B}}{2(1+H_E)}$ with $A = H_E(1+n_0) - H_O$ and $B = 4n_0\left(1 - H_E{}^2\right)$.

**Chapitre II.** Variation génétique populationnelle aux locus microsatellites

L'étude de la variation génétique aux locus microsatellites au sein et entre les populations de *L. migratoria* a été réalisée afin de tenter de répondre aux questions suivantes :

(i) A l'échelle de l'aire globale de répartition de l'espèce, les différences génétiques entre les populations corroborent-elles la classification taxonomique actuelle réalisée sur la base de critères morphométriques et comprenant onze sous-espèces (Fig. 2)?

(ii) Le patron mondial de différenciation génétique permet-il d'inférer les facteurs biotiques, abiotiques, et historiques ayant un effet majeur sur la variation génétique de l'espèce ?

(iii) A une échelle plus locale, les microsatellites sont-ils informatifs quant à l'effet potentiel des évènements de pullulation sur la variation génétique à l'intérieur et entre les populations?

Ce chapitre présente, à la suite d'une introduction méthodologique, une synthèse des résultats de ces trois questions tour à tour. Les analyses et résultats des deux premières questions, d'une part, et de la dernière question, d'autre part, sont détaillées respectivement dans les manuscrits 3 et 4, inclus à la fin du chapitre.

## 1. APPROCHE METHODOLOGIQUE

Les trois questions posées ci-dessus ont nécessité un échantillonnage combinant des populations aux propensions à pulluler et conditions environnementales différentes.

Une première raison tient au fait que les espèces présentant des fluctuations récurrentes de leurs tailles de population et/ou de leur taux de migration s'écartent probablement de l'équilibre mutation-dérive-migration, en dehors duquel il n'existe pas d'attendus théoriques clairs (*e.g.*, Leblois 2004). Il est possible de contourner cette difficulté par l'utilisation d'approches comparatives, soit d'échantillons de populations temporels collectés durant un épisode de pullulation-rémission (*e.g.*, Berthier *et al.* 2006), soit de groupes de populations caractérisés par des patrons historiques de pullulations contrastés (*e.g.*, Bailey *et al.* 2005b). Nous avons adopté cette dernière stratégie en comparant la variation génétique à l'intérieur et entre des groupes de populations de *L. migratoria* pullulantes et non pullulantes.

La nécessité de réplication, fondamentale en biologie, a également sous-tendu notre échantillonnage, afin de contrôler les effets confondants qui agissent potentiellement sur les populations naturelles et leur variation. Ce problème est particulièrement valable pour l'étude des relations de causalité entre la propension à pulluler et la composition génétique des populations. Tout d'abord, les différences de propension à pulluler entre les populations peuvent refléter des différences génétiques et/ou des différences de leurs environnements, ces derniers étant plus ou moins favorables aux augmentations de densités de populations. La question de l'adéquation entre la variation à des locus microsatellites et la variation de la propension à pulluler des populations ne peut donc être adressée correctement que si la variabilité environnementale est indirectement contrôlée sous la forme de réplicats d'échantillonnage issus d'environnements contrastés. Cette contrainte est d'autant plus importante que les populations qui partagent un même environnement, sont souvent génétiquement proches de par leur proximité géographique.

L'analyse de la variation génétique au sein et entre les populations pullulantes nécessite également des réplicats populationnels puisque, en dehors du potentiel effet des évènements de pullulation, un grand nombre de variables, géographiques, environnementales, et historiques affectent la variation génétique neutre à l'intérieur et entre les populations. Un exemple simple est celui des îles, où les populations ont généralement une variation génétique moindre que les populations continentales, de par leur histoire de fondation et les plus faibles capacités biotiques des milieux insulaires (Frankham 1997). Pour les espèces cosmopolites, la variation génétique à l'intérieur des populations peut également évoluer en fonction des fortes variations écologiques de l'aire d'habitat. Par exemple, les populations de la mouche domestique, *Musca domestica*, des latitudes les plus septentrionales subissent des évènements de fondation chaque hiver, durant lesquels la variation génétique à l'intérieur des populations est perdue par dérive (Black et Krafsur 1986). La variation génétique à l'intérieur des populations est également souvent érodée le long des voies de colonisation, du fait d'évènements successifs de fondation (*e.g.*, Estoup *et al.* 2001).

En conclusion, la validité de nos résultats d'analyse des génotypes microsatellites des populations de *L. migratoria* repose sur l'adoption d'une approche comparative prenant en compte des populations avec des propensions à pulluler et des environnements contrastés. Les Fig. 10 et 11 illustrent nos efforts d'échantillonnage en ce sens (voir aussi les manuscrits 3 et 4).

**Fig. 10. Patrons de pullulation des échantillons populationnels de *L. migratoria* génotypés aux marqueurs microsatellites.** Les patrons de pullulation ont été définis à partir de la littérature suivante : Uvarov et Hamilton 1936; Waloff 1940; Karabag 1958; Betts 1961; Wintrebert 1970; Farrow 1979; Farrow et Colles 1980; Brown 1986; Farrow 1987; Chen 1991; Hakomori et Tanaka 1992; Randriatmanantsoa 1998; Zhang et Li 1999; Lecoq et Surkino 1999; Benfekhi et al. 2002. Les encadrés comprennent les populations utilisées pour l'étude de la variation génétique des populations historiquement pullulantes (section 4 – Manuscrit 4). Une majorité de ces mêmes populations ainsi que les autres populations ont été utilisées pour l'étude de la structure génétique de l'espèce (sections 2 et 3 – Manuscrit 3).

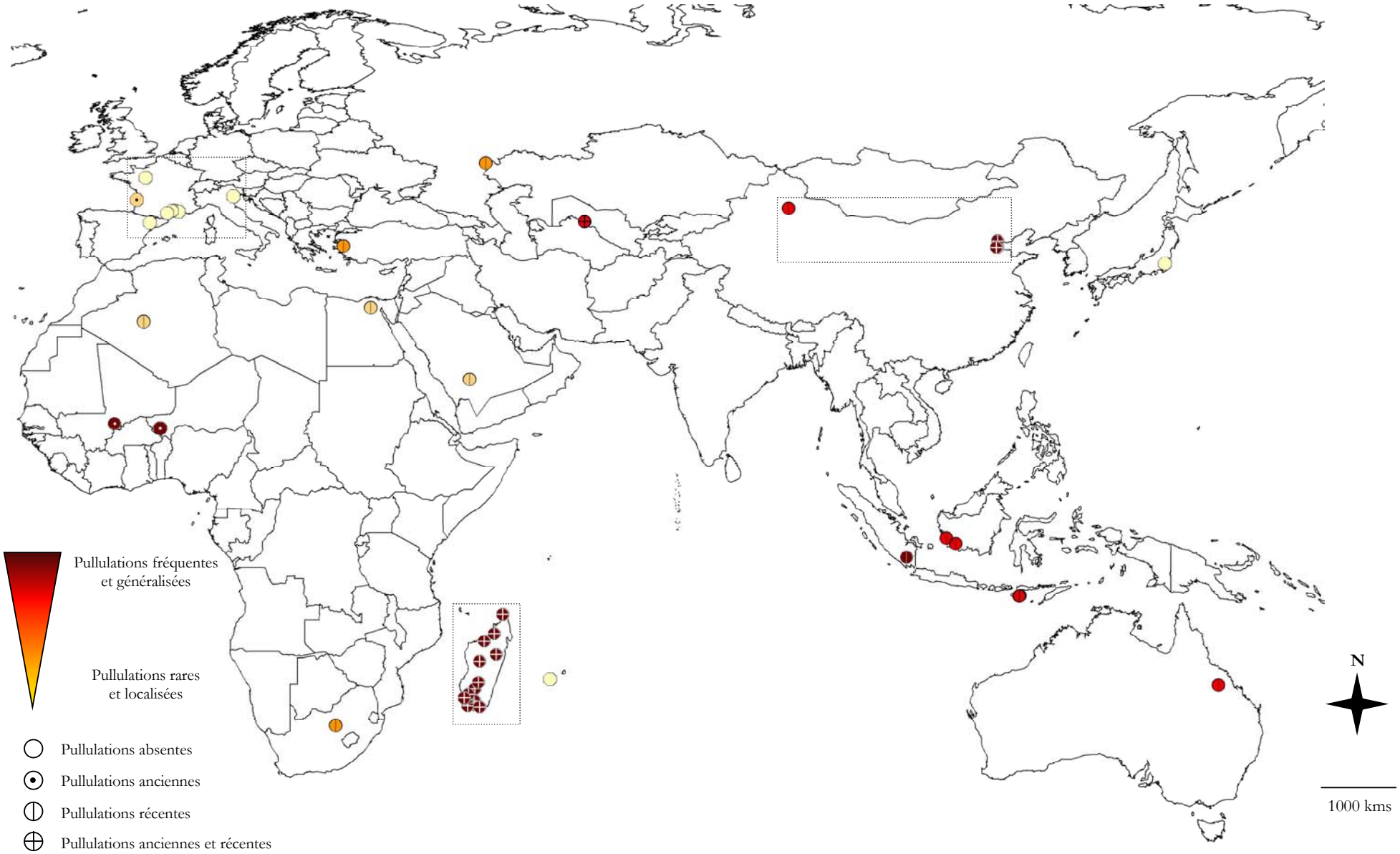
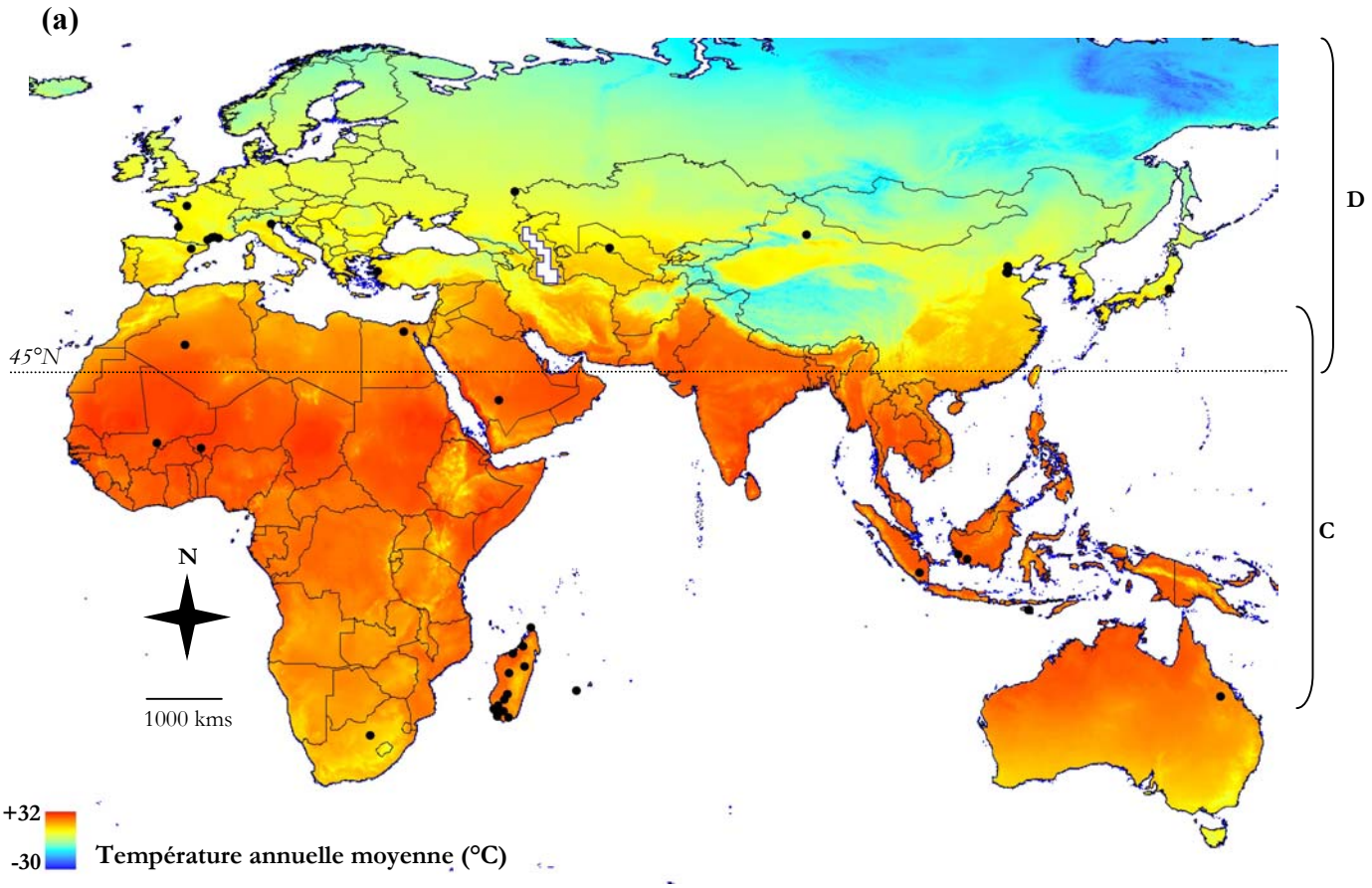
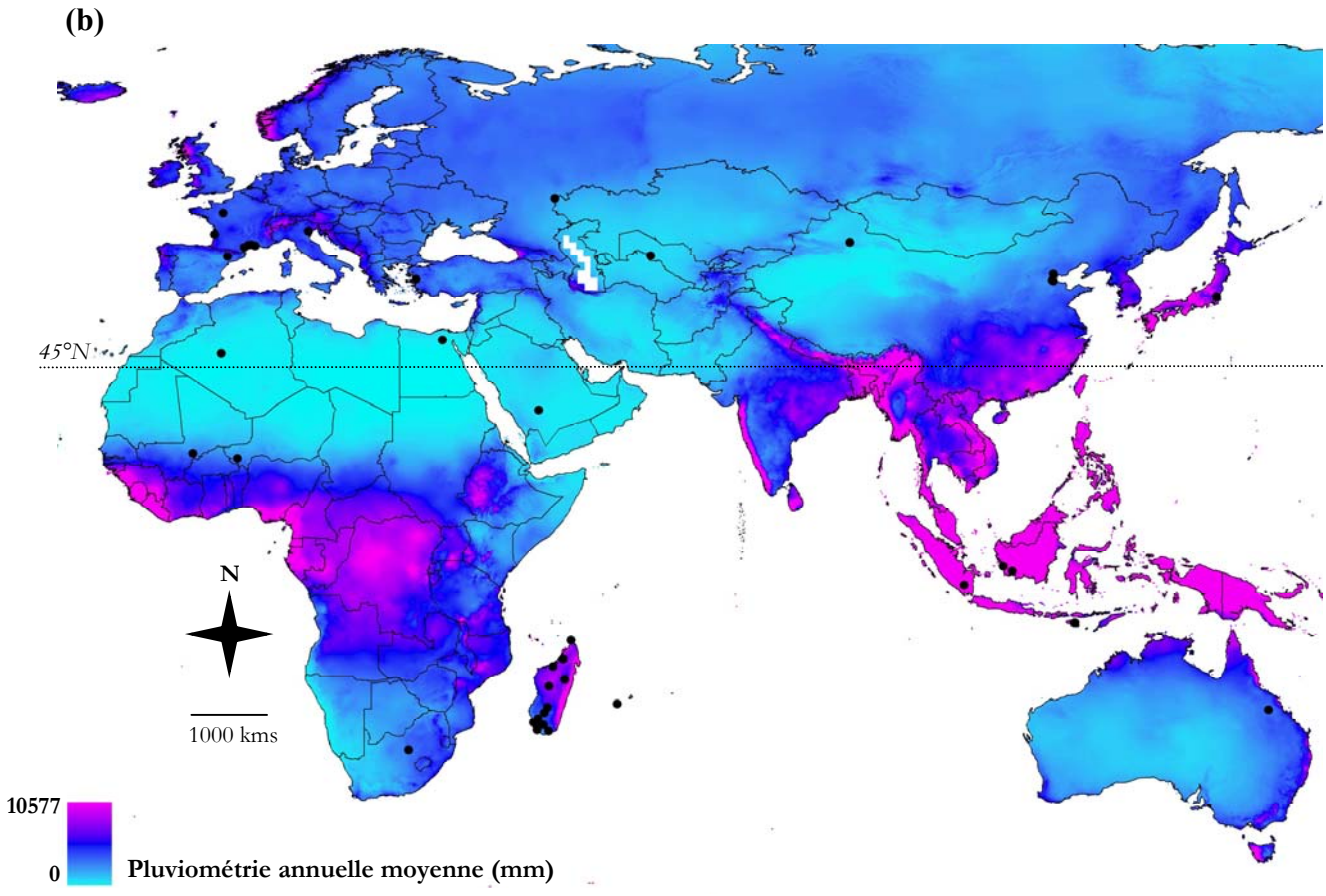


**Fig. 11. Echantillons populationnels de *L. migratoria* génotypés aux marqueurs microsatellites et variation de leurs conditions écologiques : exemples de la température, facteur déterminant de l'expression de la diapause (a) et de la pluviométrie, facteur clé dans le déterminisme des pullulations (b).** D : populations diapausant au stade œuf (univoltines ou bivoltines), C : populations à développement continu (multivoltines).



## 2. TAXONOMIE ET VARIATION GENETIQUE

Les mesures de différenciation génétique obtenues à partir des données microsatellites ont révélé de nombreuses incongruences avec la définition des onze sous-espèces réalisée à partir des données de morphométrie (Manuscrit 3). En milieu continental, les données génétiques suggèrent fortement le regroupement des sous-espèces définies par la morphométrie en deux ensembles relativement homogènes génétiquement : d'une part *L. m. migratorioides*, *L. m. cinerascens* et la sous-espèce arabe, et d'autre part les populations chinoises de *L. m. manilensis* et les sous-espèces *L. m. migratoria* et *L. m. gallica* (Fig. 2). Les sous-espèces tibétaine (*L. m. burmana*) et indienne n'ont pas pu être échantillonnées. L'homogénéité relative des populations de *L. migratoria* aux échelles eurafricaine et eurasiatique pourrait être expliqué par un niveau de flux de gènes suffisamment élevé pour effacer la plupart de la différenciation entre les populations.

En revanche, les données génétiques structurent plus fortement que l'approche taxonomique classique les populations insulaires ainsi que celles localisées en limite d'aire de répartition de l'espèce. Par exemple, les échantillons populationnels du Japon et du Nord de la France sont fortement différenciés des populations de la sous-espèce *L. m. migratoria* à laquelle les attribue la morphométrie. Similairement, la population de l'île de la Réunion est fortement différenciée des populations malgaches appartenant à la même sous-espèce *L. m. capito*. En Indonésie, les populations de Bornéo sont génétiquement distinctes des populations de Sumatra et Sumba, pourtant relativement proches géographiquement et appartenant à la même sous-espèce *L. m. manilensis*. Une exception est la population de la sous-espèce australienne génétiquement peu différenciée des populations de *L. m. manilensis* provenant de Sumatra et Sumba. Ces niveaux de différenciation relativement élevés, parfois associées à une variabilité inférieure à celles des autres populations continentales, s'explique probablement par des effets conjugués de fondation, de dérive et d'isolement génétique.

## 3. FACTEURS STRUCTURANT LA VARIATION GENETIQUE DE L'ESPECE A L'ECHELLE DE L'ENSEMBLE DE SON AIRE DE REPARTITION

La variation génétique de l'espèce semble être structurée par des facteurs géographiques, tels que les mers et les chaînes de montagnes (Manuscrit 3). Ainsi, l'ensemble des populations des îles du Pacifique et de l'Océan Indien ont montré une différenciation génétique moyenne à forte aux marqueurs microsatellites. Les deux grands groupes génétiques continentaux, regroupant les populations asiatiques et nord-européennes d'une part, et africaines et sud-européennes d'autre part, sont séparés par la chaîne de massifs montagneux traversant l'hémisphère nord des Alpes au Tibet, en passant par le Caucase.

Des facteurs écologiques semblent également structurer la variation génétique au sein de *L. migratoria*. Les données microsatellites ont mis en évidence une plus forte proximité génétique des populations orientales (Australie et Indonésie) avec les populations afrotropicales (Afrique et Madagascar) qu'avec les populations asiatiques, et ce en dépit d'un plus grand éloignement géographique. Les populations australiennes, indonésiennes, africaines, et malgaches partagent un même environnement tropical, qui contraste avec l'environnement tempéré asiatique, notamment par le rythme des saisons et l'absence d'une diapause hivernale.

Enfin, il est probable que la structure génétique observée aux marqueurs microsatellites corresponde parfois à des évènements historiques. Par exemple, la population ouest-européenne de *L. m. gallica* appartient au groupe phylogénétique asiatique fortement pullulant, et est fortement différenciée des autres populations ouest-européennes voisines et non pullulantes. Il est concevable que cette population ait été fondée au cours d'une pullulation ancienne originaire de l'Est, la dernière recensée dans cette aire datant du XIV$^{ème}$ siècle, provenant de la Mer Noire et ayant atteint l'Angleterre (Waloff 1940). En revanche, la structure génétique observée aux marqueurs microsatellites coïncide peu avec les patrons géographiques d'occurrence ou d'absence d'événements de pullulation. Ainsi, des populations génétiquement proches peuvent être caractérisées par des niveaux de pullulation récente ou ancienne très différents, et à l'inverse des populations fortement différenciées peuvent être caractérisées par des niveaux de pullulation similaires.

## 4. ECHELLE LOCALE : VARIATION GENETIQUE DES POPULATIONS EN REGIONS HISTORIQUEMENT PULLULANTES

L'étude des marqueurs microsatellites à une échelle locale s'est avérée informative quant à l'effet des évènements de pullulation sur la variation génétique à l'intérieur et entre les populations (Manuscrit 4). Pour cela, trois zones ont été étudiées plus finement, les zones historiquement pullulantes malgache et chinoise et la zone historiquement non pullulante d'Europe de l'Ouest (Fig. 9). L'hypothèse implicite ici est que les différences/similitudes observées entre les aires pullulantes chinoise et malgache d'une part et l'aire non pullulante d'Europe de l'Ouest d'autre part sont liées aux évènements de pullulation.

Les diversités génétiques à l'intérieur des populations pullulantes se sont avérées peu ou pas affectées par les augmentations récurrentes des tailles de populations typiques des zones historiquement pullulantes. Ce résultat est en accord avec l'attendu théorique qui prédit que la taille efficace populationnelle à long terme est grossièrement la moyenne harmonique de la taille au cours du temps (Motro et Thomson 1982). Dans notre cas, la variabilité génétique au sein des populations serait donc essentiellement déterminée par les tailles efficaces de rémission, d'où une similitude importante entre zones pullulantes et non pullulantes. D'autre part, la forte diversité au sein des populations indique que les locustes solitaires des périodes de rémission conservent des tailles efficaces de populations élevées. Ce résultat rend peu probable l'hypothèse de réductions drastiques de la viabilité des populations au moment du déclin démographique, comme cela a pu être suggéré pour une autre espèce de locustes *S. gregaria* (Ibrahim *et al.* 2000).

En revanche, le niveau de flux de gènes entre les populations pullulantes est environ dix fois plus élevé qu'entre les populations non-pullulantes, ce qui conduit à une forte homogénéité génétique des populations pullulantes à l'échelle de l'aire d'invasion. En accord avec ce résultat, des changements majeurs de traits de comportement de migration ont été observés avec la grégarisation chez les locustes. La variation de ces traits est susceptible d'augmenter considérablement la distance efficace de migration dans les populations grégaires en fortes densités (Ellis 1953, Uvarov 1966). Il est important de noter dès à présent que des niveaux de différenciation génétique faibles ou absents entre les populations risque de fortement limiter la

puissance des approches génétiques à inférer les sources de pullulation et déterminer les routes d'invasion (*e.g.*, Roeder *et al.* 2001).

En résumé, l'étude de la variation génétique aux locus microsatellites au sein et entre les populations de *L. migratoria* nous a permis de mettre en évidence :

- De fortes incongruences entre la structuration génétique mondiale, comprenant deux grands ensembles continentaux et des petits groupes insulaires ou en limite d'aire de répartition, et la classification taxonomique actuelle, réalisée sur la base de critères morphométriques et comprenant onze sous-espèces ;
- Une absence de structuration génétique mondiale en fonction du statut pullulant ;
- A une échelle plus locale, des tailles efficaces à long terme des populations similaires entre régions pullulante et non-pullulante, ainsi que des flux de gènes très forts en régions pullulantes et plus faibles, voire absents, en région non-pullulante.

## MANUSCRIT 3 – WORLDWIDE MICROSATELLITE GENETIC VARIATION IN *LOCUSTA MIGRATORIA*, AN OUTBREAKING SPECIES PRONE TO NULL ALLELES

Marie-Pierre Chapuis, Michel Lecoq, Anne Loiseau, Sylvain Piry, Yannis Michalakis et Arnaud Estoup

En préparation.

# WORLDWIDE MICROSATELLITE GENETIC VARIATION IN *LOCUSTA MIGRATORIA*, AN OUTBREAKING SPECIES PRONE TO NULL ALLELES

Marie-Pierre Chapuis[1,2,3], Michel Lecoq[3], Anne Loiseau[1], Sylvain Piry[1], Yannis Michalakis[2], and Arnaud Estoup[1]

[1]Centre de Biologie et de Gestion des Populations, Institut National pour la Recherche Agronomique, Campus International de Baillarguet CS 30016, 34988 Montferrier / Lez, France

[2]Génétique et Evolution des Maladies Infectieuses, UMR 274 CNRS-IRD, 911 avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France

[3]Centre de coopération internationale en recherche agronomique pour le développement, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

**Correspondence**:

Marie-Pierre CHAPUIS
Centre de Biologie et de Gestion des Populations
Institut National de la Recherche Agronomique
Campus International de Baillarguet CS 30 016
34988 Montferrier / Lez, FRANCE
Phone: +33 (0)4.99.62.33.31
Fax: +33 (0)4.99.62.33.45
E-mail: chapuimp@ensam.inra.fr

**Keywords**: locust, microsatellites, null alleles, Orthoptera, outbreak, pest, simulation, subspecies

**Running head:** Population genetics of the migratory locust

**Abstract**

Samples from 25 *Locusta migratoria* populations belonging to 9 subspecies distributed over most of the species range were scored at 14 microsatellite loci. The effects of high prevalence of null alleles in microsatellite loci of this species on all statistics and methods used to analyse genetic variation were accounted for by the way of a simulation-based study. The worldwide genetic structure among populations of *L. migratoria* was substantially incongruent with the current taxonomy in eleven subspecies based on morphometric criteria. Low levels of genetic differentiation was found among populations of *L. migratoria* from the continents. These populations were genetically clustered in two large geographical groups of populations from Eurafrica and Eurasia, respectively. The exception was the West-European area, which corresponds to limits of the species range. In contrast, most of Pacific and Indian Ocean island populations were moderately to strongly differentiated from all other populations. Average heterozygosities were significantly higher in Africa and South-Europe and significantly lower in Japan and Reunion islands than in other investigated areas of the species range. Microsatellite variation did not show clear imprints of contrasted propensities to outbreak of *L. migratoria* populations. Geographical and ecological barriers to gene flow as well as historical events better explain the observed patterns of genetic variation within and among populations at the worldwide scale.

## Introduction

The migratory locust, *Locusta migratoria*, is an agricultural pest of major importance in large areas of the ancient world. The species displays unusual outbreaks with an irregular cycle and swarming individuals typical of a high-density distinctive form, the phase *gregaria* (Uvarov 1966). Using a morphometric technique originally devised to differentiate between phases of a locust, taxonomists have described eleven geographical subspecies of *L. migratoria* over the whole species range covering all grasslands of the ancient world, from Northern Eurasia to New Zealand (reviewed in COPR 1982; Remaudière 1940; Remaudière 1947). All subspecies are distributed over large geographic areas, with the exception of two forms localized to small areas in France, *L. m. gallica* limited to the region of Bordeaux (Remaudière 1947) and the form of Palavas limited to the lagoon ecosystem of the region of Hérault (Remaudière 1940). The eleven subspecies of *L. migratoria* are globally congruent with geographical occurrences or absences of outbreak events (Farrow and Colless 1980). For instance, Zolotaresky (1929) described the subspecies *L. m. capito* for widespread swarms in Madagascar; Uvarov (1936) recognized the subspecies *L. m. manilensis* for locusts originating from extensive outbreaks areas in the Philippines, Borneo, Malaysia and China; and the subspecies *L. m. migratoria* was originally described from individuals collected during the periodic plagues in the Palaearctic (Uvarov 1921). On the opposite, individuals from more dispersed and rarely swarming populations were described in their respective areas as the form of “Palavas”, and the Arabian, Indian, and Tibetan (*L. m. burmana*) subspecies. The taxonomy in eleven subspecies of *L. migratoria* is far from being considered definitive by some authors (e.g., Farrow and Colless 1980). For instance, Uvarov (1921) and Farrow and Colless (1980) placed all locusts in two broad biogeographical and ecological units, the Palaearctic and, the Afro- and Oriental-tropics. Between these regions, locusts exhibit contrasted life histories: hibernation at the egg stage is a general rule in the Palaearctic

populations, whereas tropical and sub-tropical locusts are multivoltine, with no diapause at any stage of development. No genetic studies have been achieved so far to validate or invalidate this classification.

Microsatellite loci are widely used to assess population genetic structure and also to study evolutionary relationships below the species level (e.g., Estoup and Angers 1998). Microsatellite variation among populations over very large geographic scales has been investigated in various insect species. Examples include Diptera, such as *Drosophila species* (e.g., Pascual et al. 2001), *Ceratitis capitata* (Gasperi et al. 2002), *Musca domestica* (Krafsur et al. 2005), and *Bactrocera oleae* (Nardi et al. 2005), Hymenoptera (*Apis mellifera*; e.g. Estoup et al. 1995; Franck et al. 2001), and Lepidoptera (e.g., *Lymantria dispar*; Bogdanowicz et al. 1997). Although, microsatellite loci may be subject to size homoplasy due to convergent mutations (e.g. Estoup et al. 2002), these population genetic surveys on highly vagile, fecund, and colonizing cosmopolitan insects, showed the high efficiency of microsatellite to differentiate populations and infer groups of genetically related populations. In some insect taxa however microsatellite genotyping may be markedly error-prone given that frequent null alleles (e.g. Lepidoptera, reviewed in Meglecz et al. 2004). Orthopteran species, and in particular *L. migratoria*, seem precisely to be concerned by a high prevalence of microsatellite null alleles, since observed heterozygosities are frequently lower than expected heterozygosities at most studied microsatellite loci on this order (e.g. Hamill et al. 2006). To our knowledge, a very few microsatellite population genetics studies have been achieved in this order (but see Morgan-Richards et al. 2000 and Streiff et al. 2006). This is probably because the question of how the presence of frequent null alleles might affect most statistics and methods traditionally used to analyse genetic variation has rarely been addressed. Indeed, to date, only the impact of the presence of null alleles in genotype datasets

on population differentiation estimation and the methods available to correct for a potential bias have been recently investigated (Chapuis and Estoup 2006).

In the present study, genetic variation within and among *L. migratoria* subspecies was assessed by genotyping 14 microsatellite loci in 25 populations sampled over most of the species range. The following questions were more specifically addressed: (i) Are the levels of genetic differentiation between subspecies and populations within subspecies congruent with the current taxonomy in 11 subspecies based on morphometric criteria? (ii) Are the microsatellite markers useful to infer the biogeographic factors driving the pattern of genetic variation at the worldwide scale? (iii) Are differences in propensity to outbreak and neutral genetic variation between populations congruent? Furthermore, we used computer simulation based on the coalescent (Hudson 1990) to account for biases null alleles may introduce in microsatellite data analysis.

## Material and Methods

*Population sampling and Genotyping*

From 2001 to 2004, we collected a total of 25 population samples (19 to 32 individuals per sample) distributed in much of the distribution area of *L. migratoria* across the ancient world (Fig. 1). The sampling encompassed populations from *L. m. capito, L. m. cinerascens, L. m. manilensis, L. m. migratoria, L. m. migratorioides, L. m. gallica*, the Australian, and the Arabian subspecies and the form of 'Palavas'. The Tibetan and Indian subspecies were not sampled. Two contrasted historical patterns of *L. migratoria* outbreak events were represented in our sampling (Fig. 1). Major plagues have threatened Madagascar and China since historical times (Randriamanantsoa 19998; Zhang and Li 1999). In sub-Saharan Africa, the incidence of large spreading plagues has lasted until the last 50 years (Betts 1961). Outbreak events, albeit less widespread, has also been recurrently reported in Middle Europe (Waloff

1940). Intensive outbreaks have recently developed in Indonesia associated with agriculture development (Lecoq and Sukirno 1999). North Africa is also a newly concerned area but outbreak events remain minor and local (Benfekhi et al. 2002). In Western Europe, the last serious outbreak event, that probably originated from the Western coast of the Black Sea, dated from the XIV$^{th}$ century (Waloff 1940).

Fourteen microsatellite loci (OZC9, OZC35, OZC76 in Zhang et al. 2003 and LM1-88, LM10-78, LM2-B, LM2-Λ, LM3-Ω, LMT-113, LMT-137, LMT-177, LM10-180, LM11-121, LMT-133 in Chapuis et al. 2005) were genotyped, using fluorescent PCR and a MegaBACE sequencing machine (Amersham Biosciences), as described in Chapuis et al. (2005). OZC9 was amplified within the multiplex set 1 of Chapuis et al. (2005) with a primer concentration of 0.5 μM, OZC35 within the multiplex set 2 with a primer concentration of 0.3 μM, and OZC76 within the multiplex set 2 with a primer concentration of 0.2 μM. To minimize genotyping errors, a minimum of two replicate PCRs (Polymerase Chain Reactions) per sample per locus were conducted and alleles were scored, using the Genetic Profiler v.1.0 software (Amersham Biosciences), independently by two of the authors. Alleles included in the final consensus genotypes were observed at least twice; if observed only once, an additional replicate was conducted. We included one negative control and eight positive controls (samples with known genotypes) with every 96 PCR reactions as checks for contamination and to standardize genotypes among experiments.

*Null alleles*

Deviation from Hardy-Weinberg equilibrium (HWE) was tested for each locus separately and within each population using Fisher exact tests implemented in the software GENEPOP 3.3 (Raymond and Rousset 1995). Corrections for multiple tests were performed using the false discovery rate approach (Benjamini and Hochberg 1995). If observed genotype frequencies

deviated significantly from HWE, the program MICRO-CHECKER (van Oosterhout et al. 2004) was used to determine the most probable cause among genotyping error (e.g. short allele dominance, Wattier et al. 1998) and presence of null alleles. A maximum likelihood estimate of the frequency of null alleles (Dempster et al. 1977) was calculated using the program FreeNA (Chapuis and Estoup 2006).

Because microsatellite null alleles have high prevalence in *L. migratoria* (Chapuis et al. 2005; '*Results*' section), we used computer simulations based on the coalescent (Leblois et al. 2003) to assess the effect of the null allele presence and of the traditionally used correcting method (the so-called *INA* method) on analyses used in the present study to analyse genetic variation. We first focused our simulations on five statistics traditionally used for summarizing within-population genetic variation by assuming a single isolated population at mutation-drift equilibrium of effective size $N_e$. The Bayesian analysis implemented in the program BAPS for delineating clusters of population samples (Corander et al. 2003) was then tested assuming a migration model, in which two populations of equal effective size $N_e$ exchange migrants at a rate $m$. Finally, we evaluated the expectation of satisfactory efficiency of Neighbour-Joining algorithm (Saitou and Nei 1987) and Cavalli-Sforza and Edwards (1967) distance ($D_C$) for recovering correct tree topologies assuming a population split model. To this aim, we considered five populations assumed to evolve following the model tree specified in Figure 2, in which each population of equal effective size $N_e$ do not exchange any genes with others since the corresponding split event.

For all three population models, three types of genotype datasets were simultaneously generated. In the first, all alleles were visible. Both other datasets (same individuals) include null alleles, one being uncorrected and the other corrected for these null alleles. Mutations in the repeat region and the primer sites, leading to the presence of null alleles were simulated as described in Chapuis and Estoup (2006). In the population split model, the gene copy used for

the design of the microsatellite primers was randomly chosen in one of the five populations. Simulated genotype datasets were corrected for null alleles as described in Chapuis and Estoup (2006). We simulated genotype datasets of 14 independent loci and 60 genes for each population. The parameter $N_e\mu_R$ (with $\mu_R$ the mutation rate in the repeat region) was set to 5, as this value generates mean values of heterozygosities similar to those observed at our microsatellite markers (i.e., 0.84 ± 0.01). The tested values of the parameter $N_e\mu_B$ (with $\mu_B$ the mutation rate in the binding sites of the two primers) varied from 0.05 to 0.5 in order to explore most of the null allele frequency range.

*Genetic variation within and among populations*

For each genotyped population, we estimated over all loci the mean expected heterozygosity ($H_E$, Nei 1987) and its confidence interval, computed by bootstrap resampling over loci (10,000 replicates). Null alleles have been shown to have a limited effect on $H_E$, contrary to other statistics summarizing genetic variation within populations, such as the allelic diversity, the allelic diversity corrected for sample size, and the observed heterozygosity (see '*Results*' section). When relevant, comparisons of $H_E$ among populations sets (e.g., differing by the BAPS genetic cluster they belong to or their outbreak patterns) were performed. To this aim, we used two-sided permutation tests following Goudet et al. (2002; 10,000 iterations), but adding at each iteration a bootstrap resampling over loci to account for among-locus sampling variance (Felsenstein 1985) using a program written by M.-P. Chapuis.

For each population, genotypic differentiation over all loci was tested using Fisher exact tests (GENEPOP 3.3; Raymond and Rousset 1995). Corrections for multiple tests were performed using the false discovery rate approach (Benjamini and Hochberg 1995). The level of differentiation between populations was quantified by computing pairwise estimators of $F_{ST}$ (Weir 1996). Because all populations were shown to harbour null alleles (see '*Results*'

section), we used the so-called *ENA* correction method to efficiently correct for the positive bias induced by the presence of null alleles on $F_{ST}$ estimation (Chapuis and Estoup 2006). $F_{ST}^{\{ENA\}}$ values were computed using the package FreeNA (http://www.montpellier.inra.fr/URLB/).

We applied a Bayesian analysis implemented in the program BAPS 3.1 for estimating hidden substructure across the species range (Corander et al. 2003; 2004). The program applies a stochastic optimisation instead of a Markov Chain Monte Carlo algorithm to infer population structuring with restricted computational resources. The method determines clusters of population samples minimizing Hardy-Weinberg and linkage disequilibrium within the clusters. In this method, the number of clusters is treated as an unknown parameter. This method has been shown very conservative in identifying population structure (Waples and Gaggiotti 2006). The method hence determines the uppermost level of population structure, grouping together panmictic populations that exchange migrants at a high level. Although BAPS better discriminate different gene pools in the presence of null alleles, its power has been shown to remain low under high gene flow (see '*Results*' section).

A Neighbor-Joining tree (Saitou and Nei 1987) relating the populations was constructed based on microsatellite data using the chord-distance (Cavalli-Sforza and Edwards 1967; $D_C$). To assess the stability of the tree nodes, we ran 2000 bootstrap replications (Hedges 1992). The trees were calculated with the POPULATION software package (Olivier Langella, CNRS UPR9034, France) and graphically displayed with TREEVIEW software (Page 1996). The genetic distance of Cavalli-Sforza and Edwards (1967) was chosen because (i) it seems to be the most efficient distance for obtaining a correct tree topology (Takezaki and Nei 1996), (ii) it makes no assumption regarding constant population size or mutation rates among loci, and (iii) the null allele bias for this genetic distance is low and similar for a large range of splitting times, suggesting the tree topologies based on pairwise

$D_C$ distances to be little affected by the presence of null alleles (Chapuis and Estoup 2006). For this analysis only, we corrected our dataset following the so-called *INA* method of Chapuis and Estoup (2006) using the package FreeNA. Corrected datasets have been shown to provide more often correct tree tropologies than uncorrected ones (see '*Results*' section).

## Results

### *Null allele prevalence*

Each population sample deviated from HWE (i.e., heterozygote deficiency; $P < 0.0001$), with 5 to 13 single-locus exact tests significant after false discovery rate correction. Specimens that failed to amplify at some loci did not yield PCR products after two or three attempts of PCR reactions whereas the same DNA samples were successfully amplified at other loci. This strongly suggests the presence of null alleles for most loci and populations. In agreement with this, MICRO-CHECKER showed that the general excess of homozygotes is distributed across most allele size classes. Estimated frequencies of null alleles per locus per population ranged from 0 to 0.753, with frequencies averaged over loci varying from 0.077 to 0.250 according to the population. The mean allele frequency over all populations and loci was 0.189.

### *Genetic variation within populations*

Simulated microsatellite datasets showed that the presence of null alleles decreased estimates of all statistics summarizing genetic diversity within populations (Fig. 3). A low impact of null alleles was however observed on the expected heterozygosity (Nei 1987; $H_E$). For a null allele frequency of 0.19 (i.e. our mean value over loci and populations), values of $H_E$ and $V$ were only slightly lower to those in the absence of null alleles. Conversely, $A$, $R_S$, and $H_O$ distributions were almost separated from those in the absence of null alleles. The traditionally used correction procedure for null alleles partially rectified the bias induced by null alleles in

*A* and $H_O$, but continued to generate underestimated values. The correction procedure increased the negative bias induced by null alleles in $H_E$ and $R_S$ with estimates reaching values lower than those obtained with uncorrected datasets.

When using our real dataset, all microsatellite loci genotyped in *L. migratoria* populations were highly variable, the mean expected heterozygosity varying from 0.68 to 0.92 according to the population sample (Fig. 4). The mean expected heterozygosity was significantly larger for populations of the BAPS genetic cluster comprising African and Southern European populations (see Fig. 1 and below) than for the others populations ($P = 0.0011$). Moreover, the mean expected heterozygosity was significantly lower for the Japan and Reunion islands (populations 16 and 22) than for the other populations ($P = 0.0016$). The levels of genetic variation within-population did not differ between the two contrasted outbreak patterns ($P = 0.2962$).

*Genetic structure*

Fig. 5a showed that the Bayesian analysis for delineating clusters of groups of individuals of Corander et al. (2003) had limited power to detect panmictic populations for high levels of gene flow. For instance, the two simulated populations were discriminated in less than 50% of simulated datasets without null alleles for $N_em = 2$. This is in agreement with the results of Waples and Gaggiotti (2006), which showed that BAPS could reliably discriminate different gene pools only under restricted migration ($N_em = 1$). The presence of null alleles increased the power of BAPS to discriminate populations, especially for high levels of gene flow (Fig. 5a). The correction procedure continued to increase the power of BAPS, but in a lower extent. For instance, for $N_em = 2$, BAPS found two different populations in 18, 83, and 55 of 100 simulated files without null alleles, with 19% null alleles and uncorrected, and with 19% null alleles and corrected, respectively. The power of BAPS remained however low for high levels

of gene flow even for datasets with null alleles (e.g., two populations found in only 10 files for $N_e m = 4$).

When using our *L. migratoria* real dataset, significant genotypic differentiation between populations samples ($P \leq 0.05$) were observed in 439 of the 465 pairwise comparisons. Non significant differentiation concerned many African and all Malagasy pairwise comparisons (Table 1). Fig. 6 shows that pairwise $F_{ST}^{\{ENA\}}$ distribution was skewed, with an excess of small values (e.g., 75% of values $\leq 0.1$) and a long tail with large values (e.g., Japan vs. Reunion islands: $F_{ST}^{\{ENA\}} = 0.265$). The BAPS analysis provided the highest posterior probability for a substructure of nine population clusters (Fig. 1; probability of the model = 1). Some of the genetic clusters of populations were widely distributed geographically: one encompassed the whole African continent, the Arabian peninsula, and Southern Europe; another included Chinese, Uzbekistan, Russian and a French populations. On continental lands, two French populations, in the limits of the species range, sorted out separately in a single genetic cluster (populations 9 and 11). The other genetic clusters included populations located on insular habitats and were spatially limited to their island limits, except for the populations of Sumatra, Sumba, and Australia that formed a single genetic cluster. The congruence between traditional taxonomical determination and BAPS genetic clustering was globally poor. For instance, the genetic cluster of Eurafrican populations comprised the *L. m. migratorioides*, *L. m. cinaerescens*, and the Arabian subspecies. On the opposite, the *L. m. manilensis* subspecies was split in three distinct genetic clusters: the groups of populations from China, from Borneo, and from Sumatra, Sumba, and Australia. We also found a weak congruence between patterns of genetic variation among *L. migratoria* populations and geographical occurrences and absences of outbreak events. For instance, the populations from Spain (8) and Mali (2) present largely contrasted patterns of outbreak events (Fig. 1), though they were weakly differentiated at microsatellite markers

($F_{ST} = 0.006$) and were grouped into the same BAPS genetic cluster. Conversely, although geographically close, the populations from Borneo (17 and 18) and Sumba (20) had higher levels of genetic differentiation ($F_{ST} = 0.044$) and belonged to two distinct BAPS genetic clusters, though they shared the same high intensity and recent timing of outbreak events (Fig. 1).

*Phylogenetic relationships*

Fig. 5b shows the performance of the Neighbour-Joining algorithm and the Cavalli-Sforza and Edwards' (1967) distance to detect the correct topology in presence and in absence of null alleles. Note that topological errors did not increase linearly with splitting times both in absence and in presence of null alleles. The power to detect the correct topology was limited for small splitting times (e.g., 42% for $t$ = 200 and in presence of null alleles) and for large splitting times (e.g., 53% for $t$ = 25000 and in presence of null alleles). On the other hand, the tree topology was correctly found in most replicates for splitting times of some thousands of generations (e.g., 76% for $t$ = 1000 and in presence of null alleles). These results parallel those of Takezaki and Nei (1996), which showed that the sampling error of genetic distances is large for low divergence levels and that the average values of the $D_C$ genetic distance reach a plateau at high divergence levels. The presence of null alleles only slightly decreased the percentage of correct topologies (i.e., $P_{TREE}$; Fig. 5b). For small splitting times, the correction procedure partially rectified the errors in recovering correct topologies induced by null alleles. For large splitting times, the percentages of correct topologies reached values larger than those obtained in the absence of null alleles (Fig. 5b). Additional simulations showed that the performance of other genetic distances traditionally used for recovering tree topologies (i.e., Nei et al.' (1983) distance $D_A$, Nei's (1978) standard distance $D_S$, and

Goldstein et al. (1995) distance ($(\delta\mu)^2$) were poorer than that of the Cavalli-Sforza and Edwards' (1967) distance, in both the absence and presence of null alleles (results not shown).

When using our real *L. migratoria* microsatellite dataset corrected for null alleles, the Neighbour-Joining population tree constructed from Cavalli-Sforza and Edwards' (1967) distance differentiated two large biogeographical regions: a first group including the Eurasian and African continents, and a second group including populations of the Pacific and Indian ocean islands (Fig. 7). The relationships among populations could not be resolved further inside the continental group. This group, nevertheless, comprised populations from 7 of the 9 sampled subspecies determined by the classical taxonomical approach. In contrast, the group of island populations was genetically highly structured, with large bootstrap supports, although the classical taxonomical approach defined only four subspecies inside this group. Moreover, *Dc* genetic distances between *L. migratoria* populations were significantly not congruent with differences in propensity to outbreak (Mantel test; $P = 0.1686$). This was suggested by the continental and island gene tree groups which are both composed of populations with contrasted outbreak patterns. Note that the phylogenetic reconstruction identified genetic groups of populations largely corresponding to those obtained using BAPS and that the use of the uncorrected dataset produced identical results.

## Discussion

*Comparison of microsatellite data and morphometrical taxonomy*

We found that the worldwide genetic structure among populations of *L. migratoria* was substantially incongruent with the current taxonomy in eleven subspecies based on morphometric criteria. Low levels of genetic differentiation were found among populations of *L. migratoria* from the continents ($0.001 \leq F_{ST} \leq 0.015$ for Africa ; $-0.002 \leq F_{ST} \leq 0.011$ for Asia; $0.005 \leq F_{ST} \leq 0.047$ for Europe). The gene tree was weakly resolved for continental

populations and the BAPS clustering method determined only a few genetic clusters of continental populations that were usually geographically widespread. In particular, *L. m. migratorioides, L .m. cinerascens,* and the Arabian subspecies on the one hand, and *L. m. migratoria,* Chinese *L .m. manilensis,* and *L. m. gallica* on the other hand were genetically clustered together. The weak genetic differentiation of continental populations could be explained by a sufficiently large level of gene flow at the scale of Europe, Asia, and Africa continents to weaken differentiation among populations. Nevertheless, the Northern-Western European population (11) and the population sampled from 'Palavas' area (9) sorted out separately in the BAPS analysis with middle level of population differentiation ($0.033 \leq F_{ST} \leq 0.066$ with other European populations). Moreover, in contrast to most continental populations, Indian and Pacific island populations generally showed a strong and significant levels of differentiation. For instance, $F_{ST}$ estimates (i.e., $0.095 \leq F_{ST} \leq 0.265$), the BAPS genetic clustering, and the gene tree showed that population samples from Japan and Reunion islands were strongly differentiated from all other populations, including those belonging to the same subspecies, *L. m. migratoria* and *L. m. capito* respectively. *L. m. manilensis* populations from Borneo were separated from those originating from Sumatran, Sumban, and Australian populations with substantial genetic differentiation ($0.044 \leq F_{ST} \leq 0.072$). The relatively high level of genetic structure among populations from islands and from areas in the limit of the species range, sometimes associated with low levels of within-population genetic variation, is likely to be explained by genetic drift, foundation events, and geographical isolation.

*Biogeographic factors associated with worldwide genetic structure*

The observed *L. migratoria* population structure shows imprints of geographical barriers to gene flow. The substantial level of differentiation of island populations indicates that although

*L. migratoria* has high dispersal abilities in continental habitats, ocean and sea represent dispersal barriers for the species. The Eurafrican and Eurasian groups of population samples separated by the BAPS genetic clustering present middle levels of pairwise $F_{ST}$ values, varying from 0.021 to 0.067. These two major biogeographical groupings seem to be physically separated by the great mountain belt of the northern hemisphere (i.e., from Alps to Tibet). Unfortunately, we were unable to obtain samples from Indian and Tibetan subspecies that would have been necessary to better address this hypothesis. The Western part of the European continent comprised populations belonging to both the Eurafrican and the Eurasian BAPS genetic clusters, suggesting that this area is a contact zone between these two relatively isolated groups of populations.

Some ecological factors also appeared to structure the genetic variation among populations at the worldwide scale. Australian and Indonesian population samples were more closely related to African ($0.035 \leq F_{ST} \leq 0.097$) than Chinese ($0.085 \leq F_{ST} \leq 0.127$) populations. Farrow and Colless (1980) have ever outlined this closer proximity of Australian samples with African than Asian samples, based on morphometrics. Africa, Australia, and Indonesia share tropical environments, opposed to the Palaearctic conditions of Asia. This result hence suggests that locusts disperse more within a biome (tropical vs. temperate) than between two different biomes.

Finally, the genetic structure of *L. migratoria* was also congruent with some historical events. For instance, the West-European *L. m. gallica* population was included into the Asian and East-European BAPS genetic cluster (Fig. 1) and was clearly related to Asian and East-European populations in the Neighbour-Joining gene tree (Fig. 7). The level of differentiation of *L. m. gallica* with Asian and East-European populations was however higher ($0.028 \leq F_{ST} \leq 0.031$) than that among Asian and East-European populations, suggesting that this population has been isolated from Asian and East-European populations since its

foundation for a substantial time. It is conceivable that this population was founded during the last Western European outbreak event that occurred at the XIV$^{th}$ century and originated from the Western coast of the Black Sea (Waloff 1940).

*Incongruence between worldwide and outbreak patterns*

The plausibility of a congruence between differences in propensity to outbreak and genetic differentiation between *L. migratoria* populations relies on the hypothesis that the outbreak events in this species are not strictly determine by environmental factors but also by population traits under genetic variation. Recently, Chapuis et al. (in prep.) showed evidence, on a morphological and behavioural based-study, that a genetically-based adaptive variation for propensity to gregarise occurs in *L. migratoria*. However, the population structure of neutral markers, such as microsatellites, do not necessarily correspond to that at loci under selection. This could explain why we did not find a congruence between patterns of genetic variation among *L. migratoria* populations and geographical occurrences and absences of outbreak events. For instance, African and European areas were genetically close (e.g. belonged to the same BAPS genetic cluster) but did not share the same patterns of outbreak events. Conversely, the geographically close populations from Borneo and Sumba belonged to two distinct BAPS genetic clusters, though they shared the same high intensity and recent timing of outbreak events.

Our results hence differ from those of Bailey et al. (2005) who showed that the non-outbreaking populations of a wingless cricket, *Anabrus simplex*, in the eastern slope of the Rockies mountains are genetically distinct from western populations, that are primarily outbreaking. However, this congruence between homogeneous units for genes and for patterns of outbreak events in this species could be explained, at least partly, by the presence of a geographical barrier to gene flow (i.e., Rockies mountains) and contrasted environments

between the two groups. The authors hence could not thoroughly test whether differences in propensity to outbreak between *A. simplex* populations may correlate with genetic differentiation among populations, independently of geographical or environmental proximity. By analysing a large number of populations with different propensities to outbreak and living in a large variety of environments, our study on *L. migratoria* is likely to represent a better way of assessing the relationship between genetic variation and outbreak pattern.

*Analysis of datasets with high prevalence of null alleles*

Heterozygote deficits were frequent at many loci in all 25 studied populations of *L. migratoria*, and the presence of null alleles was strongly support as the cause of these observed heterozygote deficits. There was hence a great need for a simulation-based study on the effect of presence of null alleles in microsatellite data analysis of this species. Because microsatellite null alleles have been found in a wide range of taxa (Dakin and Avise 2004) and additional simulations showed that our results hold for a large range of levels of genetic diversity (results not shown), conclusions of this study are of general interest in evolutionary and population genetics studies. Computer simulations showed that the presence of null alleles led to an underestimation of all statistics traditionally used to summarize genetic variation within populations, but that this bias was low for the expected heterozygosity (Nei 1987). This was somewhat expected because analyses based on second-order terms (such expected heterozygosity) are not as affected by errors as those that are based on first-order terms (for example observed heterozygosity or numbers of alleles). Moreover, the *INA* conventional method for correcting genotype data containing null alleles performed poorly. It is not surprising, since it makes the unrealistic assumption of a single null allele state whatever the null allele frequency (Chapuis and Estoup 2006). Therefore, although the bias induced on HWE is rectified, the bias induced on the level of genetic diversity remains.

Our simulations also showed that the true presence of population structure was more frequently detected by the Bayesian analysis of Corander et al. (2003) when datasets harboured null alleles, at least in cases of high levels of gene flow ($N_e m \geq 1$). This result, at first sight surprising, is consistent with the presence of null alleles leading to an upwards bias in some commonly used estimators of genetic differentiation (i.e., $F_{ST}$ and genetic distances; Chapuis and Estoup 2006). The traditionally used correction procedure for null alleles partly decreased the upwards power of BAPS to discriminate populations in the presence of null alleles. This result is consistent with the correction method decreasing the bias induced by null alleles in $F_{ST}$, but still generating overestimated values for high levels of gene flow (Chapuis and Estoup 2006).

Finally, we here directly evaluated and confirmed the expectations of Chapuis and Estoup (2006) (i) of satisfactory efficiency of the Neighbour-Joining algorithm and $D_C$ genetic distance and (ii) of better efficiency of corrected datasets than uncorrected datasets to recover the correct tree topology in the presence of null alleles. It is worth noting that datasets corrected for null alleles were even better in recovering the correct tree topology than datasets in absence of null alleles for large splitting times. Chapuis and Estoup (2006) showed that genetic distances calculated from corrected datasets were underestimated, at least when null allele frequencies were high. It is hence plausible that the correction method sufficiently underestimated large genetic distances to move them far from the plateau where the power to detect the correct topology is limited.

**Acknowledgments**

We are grateful to S. Arsol, Mr. Badaruddin, L. Baoping, D. Brown, N. Cliquennois, O. Deveci, A. El-Heneidy, P. Fontana, A. Foucart, A. Guendouz-Benrima, O. Hasbullah, Y. Hidayat, D. Hunter, Mr. Kadiran, I. Kamal, Mr. Karsono, A. Khan, P. Meunier, O.N.

Naumovich, D. Llucià Pomares, F. Ramli, T. Rachadi, G. Reynaud, X. Shengquan, S. Soumavro, Mr. Sumarli, F. Tami, S. Tanaka, S. Yskak, T. Yulianti, S. Wangpeng, Mr. Widji, Mr. Winarto, D.-X. Zhang, G. Zhiyong for providing us with *L. migratoria* population samples. M.P.C. was supported by a grant from the CNRS. This work was partly funded by INRA (scientific SPE department) and by CIRAD. YM acknowledges support from CNRS and IRD.

**Literature cited**

Bailey NW, Gwynne DT, Ritchie MG (2005) Are solitary and gregarious Mormon crickets (*Anabrus simplex*, Orthoptera, Tettigoniidae) genetically distinct? *Heredity* **95**, 166-173.

Benfekih L, Chara B, Doumandji-Mitiche B (2002) Influence of anthropogenic impact on the habitats and swarming risks of *Dociostaurus maroccanus* and *Locusta migratoria* (Orthoptera, Acrididae) in the Algerian Sahara and the semiarid zone. *Journal of Orthoptera Research* **11**, 243-250.

Benjamin Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289-300.

Betts E (1961) Outbreaks of the African migratory locust (*Locusta migratoria migratorioides* R. & F.) since 1871. *Anti-Locust Memoir* **6**, 1-25.

Bogdanowicz SM, Mastro VC, Prasher DC, Harrison RG (1997) Microsatellite DNA variation among asian and north american gypsy moths (Lepidoptera: Lymantriidae). *Annals of the Entomological Society of America* **90**, 768-775.

Brown HD (1986) New locust problem. *Antenna* **10**, 11-13.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Chapuis M-P, Estoup A (2006) Microsatellite null alleles and estimation of population differentiation. *Molecular and Biological Evolution* **in press**.

Chapuis M-P, Estoup A, Augé-Sabatier A, *et al.* (in prep.) Genetic variation for prental effects on propensity to gregarise in *Locusta migratoria*.

Chapuis M-P, Loiseau A, Michalakis Y, Lecoq M, Estoup A (2005) Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*. *Molecular Ecology Notes* **5**, 554-557.

Chen Y-L (1991) The migratory locust, *Locusta migratoria*, and its asiatic subspecies. *Ortopterists' Society & Lyman Entomological Museum, Québec*.

COPR (1982) The locust and grasshopper agricultural manual. *Centre for Overseas Pest Research, London.*

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**, 2363-2369.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1-38.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. *in Advances in Molecular Ecology. eds. Carvalho G. (Amsterdam NATO press)*, 55-86.

Estoup A, Garnery L, Solignac M, Cornuet JM (1995) Microsatellite variation in honey bee (*Apis mellifera*) populations - Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* **140**, 679-695.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consquences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.

Farrow RA (1979) Causes of recent changes in the distribution and abundance of the migratory locust (*Locusta migratoria* L.) in Australia in relation to plagues. *CSIRO Division of Entomology Commonwealth Scientific and Industrial Organization, Canberra* **9**.

Farrow RA (1987) Effects of changing land use on outbreaks of tropical migratory locust, *Locusta migratoria migratorioides* (R. & F.). *Insect Scientific Application* **8**, 969-975.

Farrow RA, Colless DH (1980) Analysis of the Interrelationships of geographical races of *Locusta migratoria* (Linnaeus) (Orthoptera : Acrididae) by numerical taxonomy, with special reference to sub-speciation in the tropics and affinities of the australian race. *Acrida* **9**, 77-99.

Franck P, Garnery L, Loiseau A, *et al.* (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity* **86**, 420-430.

Gasperi G, Bonizzoni M, Gomulski LM, *et al.* (2002) Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*. *Genetica* **116**, 125-136.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences USA* **92**, 6723-6727.

Hakomori T, Tanaka S (1992) Genetic control of diapause and other developmental traits in Japanese strains of the migratory locust, *Locusta migratoria L.* : univoltine vs. bivoltine. *Japanese Journal of Entomology* **60**, 319-328.

Hedges SB (1992) The number of replications needed for accurate estimation of the bootstrap P-value in phylogenetic studies. *Molecular and Biological Evolution* **9**, 366-369.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computation and Graphical Statistics* **5**, 299-314.

Karabag T (1958) The Orthoptera Fauna of Turkey. A synonymic and distributional Catalogue of Turkish Orthoptera. *Ankara University Sciences Faculty, Ankara, Turkey.*

Krafsur ES, Cummings MA, Endsley MA, Marquez JG, Nason JD (2005) Geographic differentiation in the house fly estimated by microsatellite and mitochondrial variation. *Journal of Heredity* **96**, 502-512.

Lecoq M, Sukirno (1999) Drought and exceptional outbreak of the oriental migratory locust, *Locusta migratoria manilensis* (meyen 1835) in Indonesia (Orthoptera: Acrididae). *Journal of Orthoptera Research* **8**, 153-161.

Loader CR (1996) Local likelihood density estimation. *Annals of Statistics* **24**, 1602-1618.

Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology* **14**, 2729-2738.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* **19**, 153-170.

Page RDM (1996) TreeView: An application to display phylogenetic trees on personal computers. *Computer applications in the biosciences* **12**, 357-358.

Pascual M, Aquadro CF, Soto V, Serra L (2001) Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. *Molecular Biology and Evolution* **18**, 731-740.

Randriatmanantsoa M (1998) Manuel sur la lutte antiacridienne. *Direction de la protection des végétaux et Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, Antananarivo*.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Heredity* **86**, 248-249.

Remaudière G (1940) Contribution à l'étude des *Locusta migratoria migratoria* L. ph. solitaria de la région de Palavas (Hérault) 1ère partie. *Revue de Pathologie Végétale et d'Entomologie Agricole de France* **17**, 147-163.

Remaudière G (1947) Sur l'existence en France d'une nouvelle sous-espèce de *Locusta migratoria* L. *Comptes Rendus des Séances de l'Académie des Sciences* **225**, 1025-1026.

Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Mathematical Biosciences* **53**, 131-147.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstruction of phylogenetic trees. *Molecular and Biological Evolution* **4**, 406-425.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389-399.

Uvarov BP (1921) A revision of the genus *Locusta* L, with a new theory as to the periodicity and migrations of locusts. *Bulletin of Entomological Research* **12**, 135-163.

Uvarov BP (1936) The oriental migratory locust (*Locusta migratoria manilensis* Meyen 1835). *Bulletin of Entomological Research* **27**, 91-104.

Uvarov BP (1966) Grass-hoppers and Locusts. *Cambridge University Press* **Volume 1**.

Uvarov BP, Hamilton AG (1936) Phase variation and rate of development in the algerian race of the migratory locust (Locusta migratoria, L.). *Bulletin of Entomological Research* **27(1)**, 87-90.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Waloff ZV (1940) The distribution and migrations of *Locusta* in europe. *Bulletin of Entomological Research* **31**, 211-246.

Waples RS, Gaggiotti OE (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **in press**.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology* **7**, 1569-1573.

Weir BS (1996) Genetic Data Analysis II. *Sinauer: Sunderland, MA.*

Wintrebert D (1970) Identité, écologie et comportement du criquet migrateur dans le sud-ouest Malgache. *Annales de la Société Entomologique de France* **6**, 35-152.

Zhang DX, Yan LN, Ji YJ, *et al.* (2003) Isolation, characterization and cross-species amplification of eight microsatellite DNA loci in the migratory locust (*Locusta migratoria*). *Molecular Ecology Notes* **3**, 483-486.

Zhang Z, Li D (1999) A possible relationship between outbreaks of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China and the El Nino episodes. *Ecological Research* **14**, 267-270.

Zolotarevsky BN (1929) Le criquet migrateur (*Locusta migratoria capito* Sauss.) à Madagascar. *Annales des Epiphyties* **15**, 185-240.

**Figure legends**

**FIG. 1. Species range of *L. migratoria* with geographical origins, taxonomical units, outbreak patterns, and genetic clustering of sampled populations**. Sampling sites with similar colors belong to the same genetic cluster as assessed by the Bayesian method of Corander et al. (2003). *Au*, Australian subspecies; *Ar*, Arabian subspecies; *Ca*, *L. m. capito*; *Ci*, *L. m. cinaerescens*; *Ga*, *L. m. gallica*; *Ma*, *L. m. manilensis*; *Md*, *L. m. migratorioides; Mi*, *L. m. migratoria*; *Pa*, 'Palavas' form. Outbreak events recorded are uncommon and local (small circle) or frequent and widespread (large circle). The outbreak pattern of sampled populations were assessed using: [1] Brown 1986, Farrow 1987, and D. Brown (Plant protection Research Institute, South Africa, pers. comm.); [2,3] Betts 1961; [4] Farrow and Colles 1980; [5] Uvarov and Hamilton 1936 and Benfekhi et al. 2002; [6] ?; [7] Karabag 1958; [8to13] Waloff 1940; [14] Chen 1991; [15] Chen 1991 and Zhang and Li 1999; [16] Hakomori and Tanaka 1992; [17to20] Lecoq and Surkino 1999; [21] Farrow 1979; [22] Wintrebert 1970; [23to25] Randriamanantsoa 1998.

**FIG. 2. Population split model for the simulation-based study of the effects of microsatellite null alleles on Neighbour-Joining tree reconstruction.** Populations do not exchange migrants after splitting. *t* was fixed to 200; 1,000; 5,000; or 25,000 generations.

**FIG. 3. Effects of null alleles and efficiency of the correction method on statistics describing within-population genetic variation.** Over 1,000 replicates, the simulated null allele frequencies were grouped into classes of 0.1 units. For each class, allelic diversity (*A*), allelic diversity corrected for sub-samples of 20 genes ($R_S$), expected heterozygosity ($H_E$; Nei 1987), observed heterozygosity ($H_O$), and allele size variance in base pairs (*V*) were computed and averaged for both uncorrected and INA corrected datasets (dotted lines). *V* was not computed for corrected datasets because the size of the null allele after correction is

unknown and arbitrarily fixed. Only genotype datasets with at least 20 genes with visible states were analysed (a number close to the smallest number of genes successfully genotyped in our samples).

**FIG. 4. Estimated values and 95% confidence intervals of expected heterozygosity (Nei 1987).** Sampling numbers with similar colors belong to the same genetic cluster as assessed by the Bayesian method of Corander et al. (2003) (see Fig. 1). Outbreak events recorded are uncommon and local (small circle) or frequent and widespread (large circle). 95% confidence interval were computed by bootstrap resampling over loci (10,000 replicates).

**FIG. 5. Effect of null alleles on BAPS and NJ clustering methods.** 100 simulated tripletwise genotype datasets (one triplet includes one file without null alleles, and both uncorrected and corrected same files with null alleles) were generated for each value of the effective number of migrants by generation ($N_em$) and of the divergence time since the ancestral split event ($t$; Fig. 2). For simulated datasets with null alleles, null allele frequency $r = 0.19 \pm 0.01$ (i.e., a value close to the mean value over all loci and populations of our real dataset). $P_{BAPS}$, percentage of replicates in which BAPS inferred the correct number of two populations; $P_{TREE}$, percentage of replicates in which the correct topology was obtained. The topology of a reconstructed tree is correct when the Robinson and Foulds' (1981) distortion index takes a value of zero. Because most evolutionists are interested in making a rooted tree, we measured topological errors by rooting trees on population 1 (see Fig. 2). The topological errors for unrooted trees are equal to or smaller than those for rooted trees, because in rooted trees an additional error may be generated in the process of putting the root (Nei et al. 1983).

**FIG. 6. Distribution of pairwise $F_{ST}^{\{ENA\}}$ values computed between 25 *L. migratoria* populations genotyped at 14 microsatellite loci.** For the density estimation of pairwise $F_{ST}^{\{ENA\}}$ values, we used the locfit function (Loader 1996) implemented in version 2.1.1 of the R package (Ihaka and Gentleman 1996; http://cran.r-project.org).

**FIG. 7. Neighbour-joining tree of 25 *L. migratoria* populations based on Cavalli-Sforza and Edwards' (1967) distances estimated at 14 microsatellite loci.** It is worth noting that we constructed an unrooted tree from our real *L. migratoria* microsatellite dataset (see Fig. 7). This is because all microsatellite loci developed for this species did not show cross-species applicability (Zhang et al. 2003; Chapuis et al. 2005), rendering impossible the genotyping of outgroup species to root our tree. Bootstrap support values are shown only where the support exceeded 50% (from 1,000 resamplings). Sampling numbers with similar colors belong to the same genetic cluster as assessed by the Bayesian method of Corander et al. (2003) (see Fig.1). Outbreak events recorded are uncommon and local (small circle) or frequent and widespread (large circle). *Au*, Australian subspecies; *Ar*, Arabian subspecies; *Ca*, *L. m. capito*; *Ci*, *L. m. cinaerescens*; *Ga*, *L. m. gallica*; *Ma*, *L. m. manilensis*; *Md*, *L. m. migratorioides; Mi*, *L. m. migratoria*; *Pa*, 'Palavas' form.

**FIG. 1**

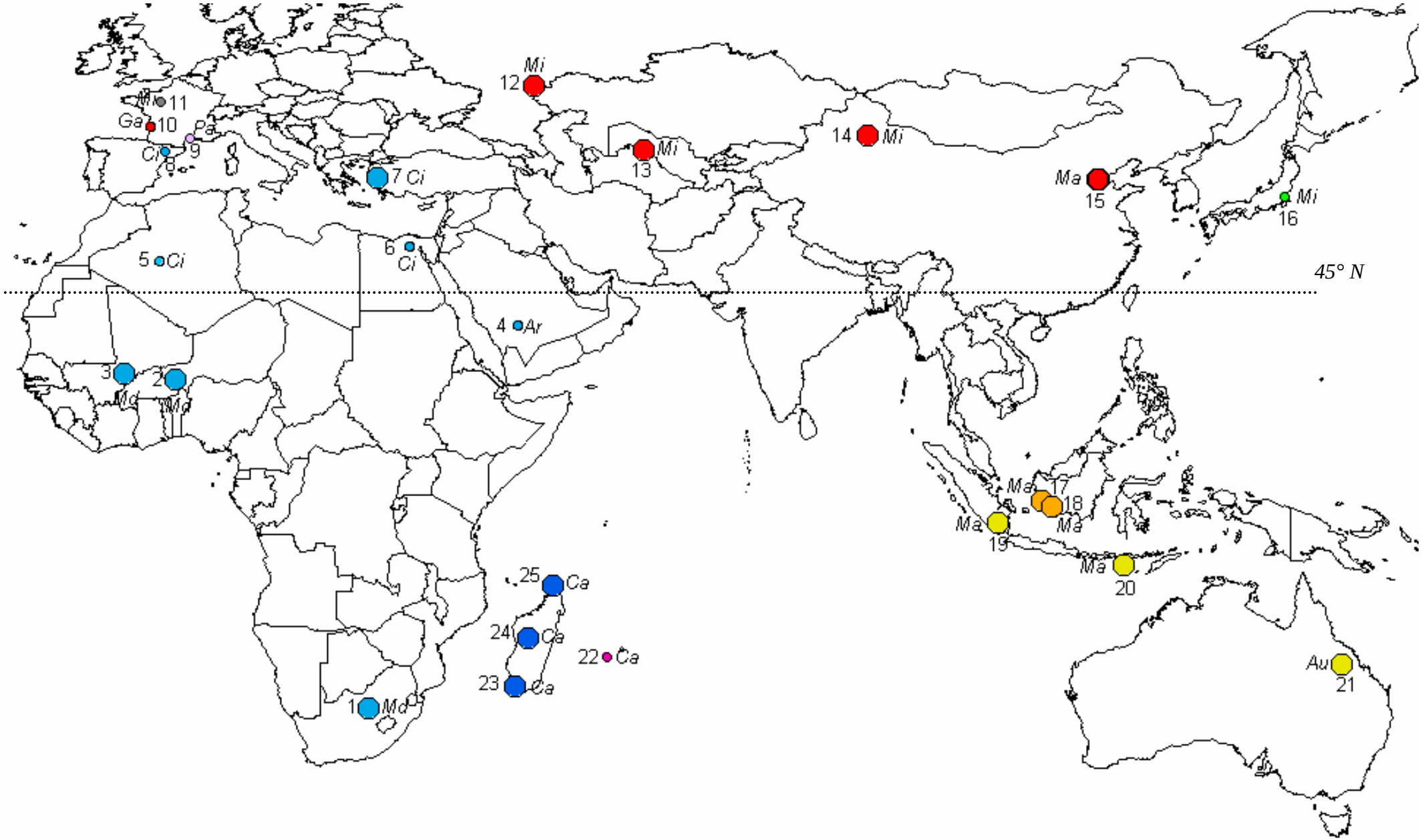

12 Mi
Mi 11
Ga 10
Pa
Ci 8
9
7 Ci
13 Mi
14 Mi
Ma 15
16 Mi
5 Ci
6 Ci
45° N
4 Ar
3 Md
2 Md
Ma 17
18 Ma
Ma 19
Ma 20
25 Ca
24 Ca
22 Ca
23 Ca
1 Md
Au 21

**FIG. 2.**

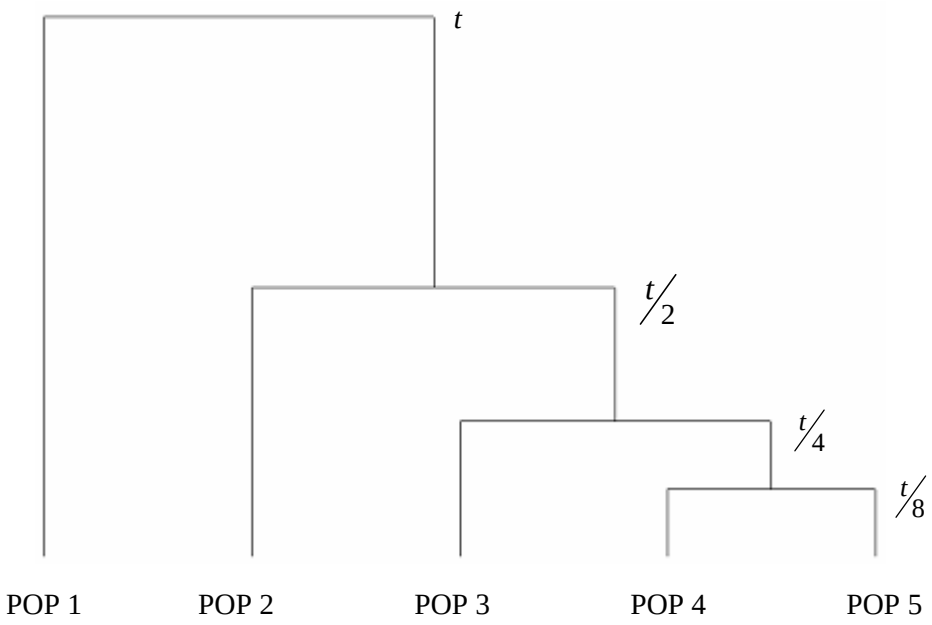


**FIG. 3.**

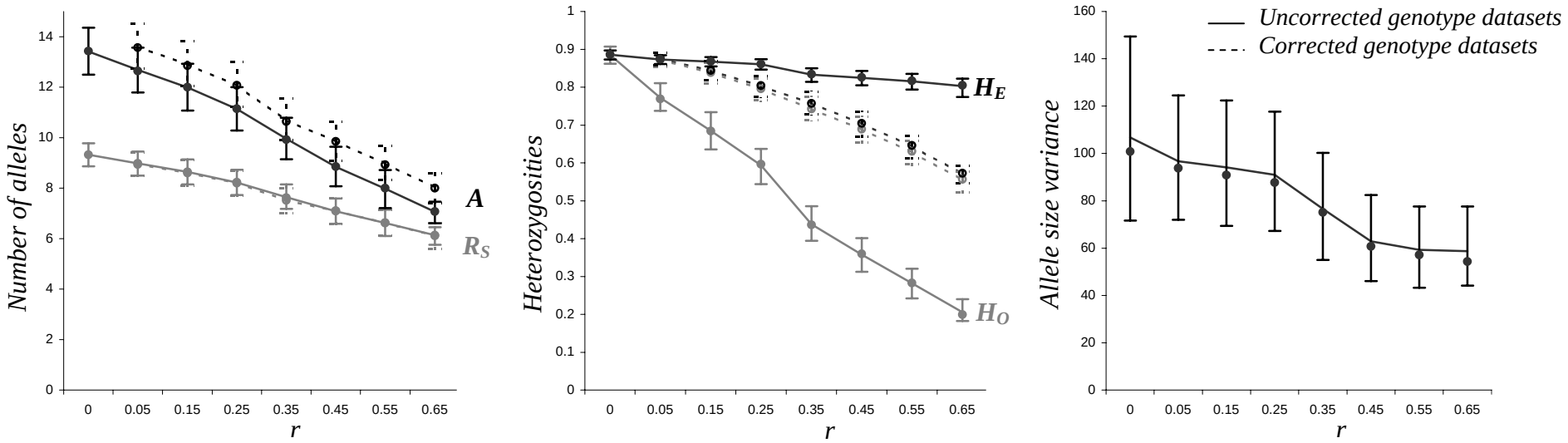

Number of alleles
A
$R_S$
r
Heterozygosities
$H_E$
$H_O$
r
Allele size variance
Uncorrected genotype datasets
Corrected genotype datasets
r

**FIG. 4**

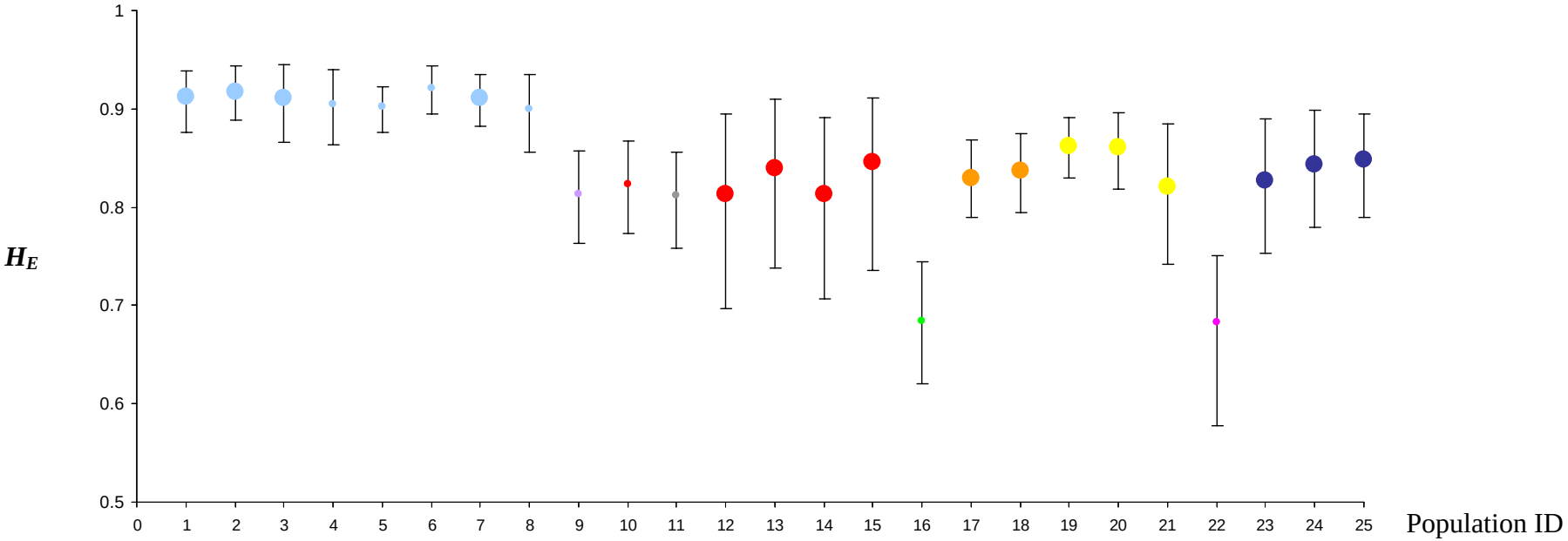
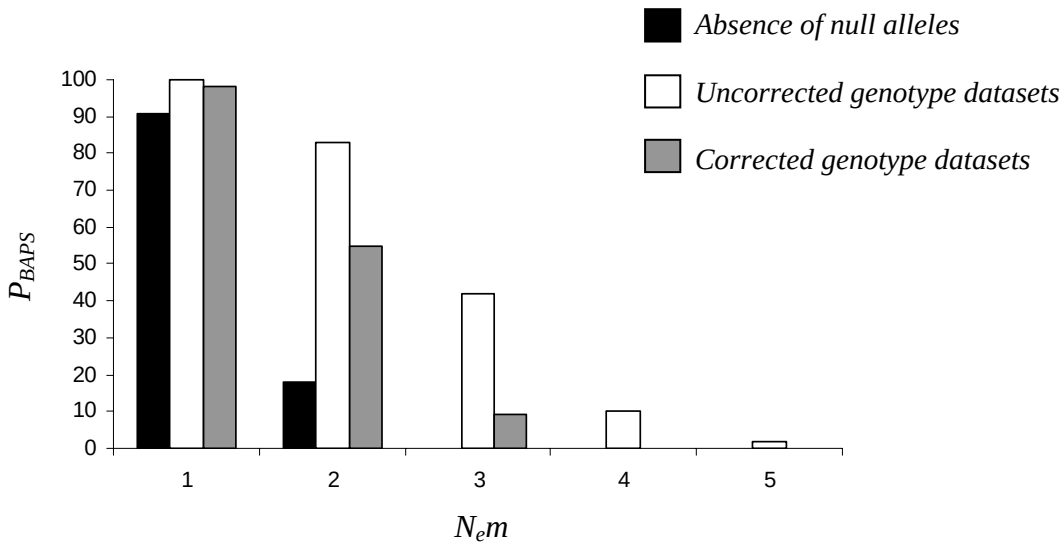

$H_E$
1
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Population ID

**FIG. 5.**

**(a)** BAPS CLUSTERING METHOD




**(b)** NJ TREE TOPOLOGY

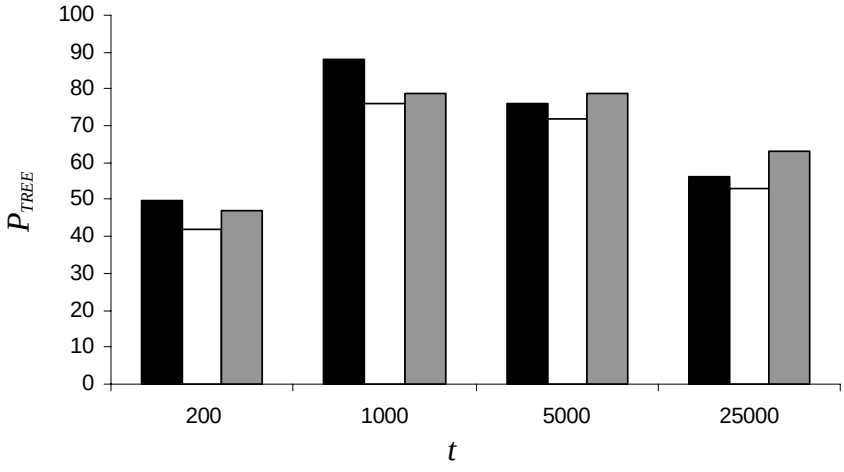


**FIG. 6**

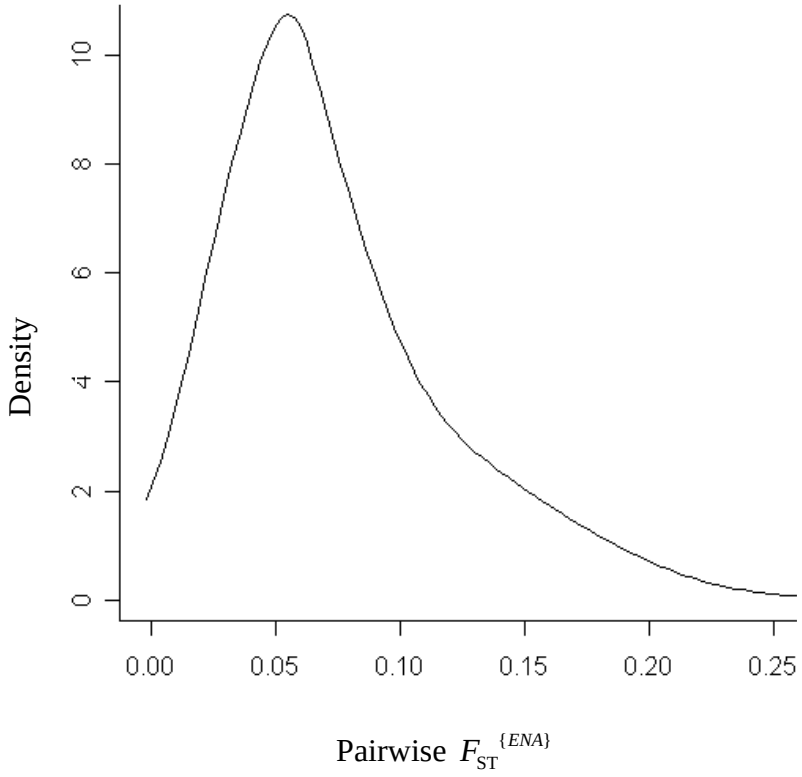
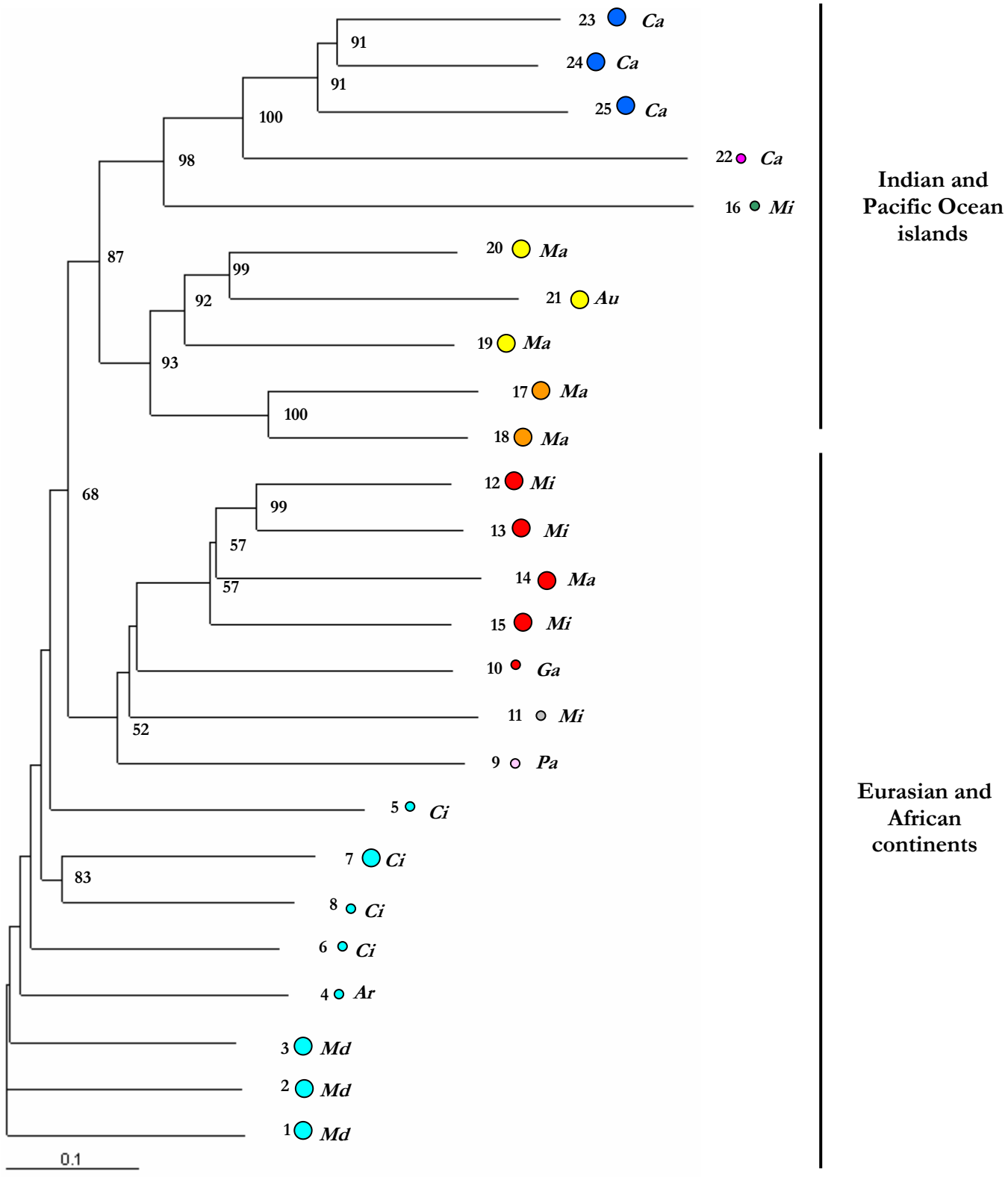


**FIG. 7**


23 Ca
91
24 Ca
91
25 Ca
100
22 Ca
98
16 Mi
20 Ma
87
99
21 Au
92
19 Ma
93
17 Ma
100
18 Ma
12 Mi
68
99
13 Mi
57
14 Ma
57
15 Mi
10 Ga
11 Mi
52
9 Pa
5 Ci
7 Ci
83
8 Ci
6 Ci
4 Ar
3 Md
2 Md
1 Md
0.1
Indian and Pacific Ocean islands
Eurasian and African continents

## MANUSCRIT 4 – CONTRASTING PATTERNS OF GENE FLOW IN OUTBREAKING AND NON-OUTBREAKING POPULATIONS OF THE MIGRATORY LOCUST, *LOCUSTA MIGRATORIA*

Marie-Pierre Chapuis, Anne Loiseau, Michel Lecoq, Alex Franc, Yannis Michalakis et Arnaud Estoup

En préparation.

# CONTRASTING PATTERNS OF GENE FLOW IN OUTBREAKING AND NON-OUTBREAKING POPULATIONS OF THE MIGRATORY LOCUST, *LOCUSTA MIGRATORIA*

Marie-Pierre Chapuis[1,2,3], Anne Loiseau[1], Michel Lecoq[3], Alex Franc[3], Yannis Michalakis[2] and Arnaud Estoup[1]

[1]Centre de Biologie et de Gestion des Populations, Institut National pour la Recherche Agronomique, Campus International de Baillarguet CS 30016, 34988 Montferrier / Lez, France

[2]Génétique et Evolution des Maladies Infectieuses, UMR 274 CNRS-IRD, 911 avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France

[3]Centre de coopération internationale en recherche agronomique pour le développement, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

**Correspondence**:

Marie-Pierre CHAPUIS
Centre de Biologie et de Gestion des Populations
Institut National de la Recherche Agronomique
Campus International de Baillarguet CS 30 016
34988 Montferrier / Lez, FRANCE
Phone: +33 (0)4.99.62.33.31
Fax: +33 (0)4.99.62.33.45
E-mail: chapuimp@ensam.inra.fr

**Keywords**: gene flow, locust, microsatellites, null alleles, Orthoptera, outbreak, pest

**Running head:** Genetic variation in the migratory locust

**Abstract**

The potential effect of outbreak events on pest species genetic variation has been rarely assessed. A major drawback in such assessment is the lack of clear theoretical expectations for populations that are likely to greatly depart from mutation-drift-migration equilibrium. Here, we circumvented this difficulty by comparing patterns of genetic variation in different sets of historically outbreaking and non-outbreaking populations, for an agricultural pest of major importance, the migratory locust, *Locusta migratoria*. We analysed genetic variation within and between 24 populations at 14 polymorphic microsatellites in Western Europe, where only ancient and low intensity outbreaks have been reported, and in Madagascar and Northern China, where frequent and intense outbreak events have been recorded over the last century. Our microsatellite survey provides strong evidence that (i) solitarious locusts persist in similarly large effective population size in both outbreaking and non-outbreaking areas, (ii) gene flow among outbreaking populations is an order of magnitude greater than among non-outbreaking populations, and (iii) historically non-outbreaking populations in Western Europe comprised a group of populations connected by gene flow, that belongs to the subspecies *L. m. cinaerescens*, and a set of isolated populations, each corresponding to a different taxonomical unit. We discussed the evolutionary implications and consequences for management strategies of the observed patterns of genetic variation.

## Introduction

Species fluctuating in their population sizes are a major concern to agriculture (e.g. some cyclic microtine rodents; swarming locusts) and human or animal health (e.g. epizootic parasites), particularly in developing countries. Thomas (1999) urged the need for management strategies of both pests and diseases to adopt an ecological approach built around a fundamental understanding of the population biology of concerned species. In particular, the design of preventive management strategies, that would reduce environmental and monetary costs, requires a clear understanding of the spatial dynamics and structure of threatening species (Hunter 2004). The study of the population dynamics and structure of threatening species can be address using direct ecological approaches, such as density estimates survey or mark-release-recapture method. Intensive field density estimates study is extremely useful to infer changes in population density and correlate them to ecological factors (Dempster 1963), but cannot be applied to assess precisely patterns of movements. Implementation of mark-release-recapture often failed for fluctuating species due to the great dispersal ability (e.g., swarming locusts; Duranton et al. 1979) and/or the low individual numbers during the remission period (e.g., cyclic microtine rodents; Wassenaar and Hobson 1998) of the concerned species. In these situations, population genetics approaches may represent an useful alternative.

Simulation and analytical studies have shown that temporal heterogeneities in demographic parameters, such as effective population size and migration rate, may have large effects on genetic variation within and between populations (Whitlock 1992; Ehrich and Jorde 2005). The sensitivity of statistics summarizing genetic variation to demographic fluctuations depends on dispersal characteristics, effective population sizes, the strength of demographic fluctuations, and the properties of the genetic markers (Slatkin 1994; Rousset 2003; Leblois 2004). Clear theoretical expectations when the organisms display demographic patterns that

greatly depart from mutation-drift-migration equilibrium are still lacking (Leblois 2004). Because species showing recurrent fluctuations in their population sizes are unlikely to be at mutation-drift-migration equilibrium, interpreting measurements of genetic variation and its distribution in terms of biological parameters such as migration rates or effective population sizes may be difficult. This difficulty might be at least partly circumvented by using comparative approaches either among temporal samples of a population set collected during a period of variation in population size and/or migration rate, or among sampled population sets with contrasted historical patterns of demographic fluctuations. To our knowledge, such comparative approaches have been rarely achieved (but see Berthier et al. 2006 for a temporal microsatellite survey).

The migratory locust, *Locusta migratoria*, provides an ideal biological model to compare the population dynamics of historically fluctuating and non-fluctuating populations using a population genetic approach. *L. migratoria* is an agricultural pest of major importance in large areas of the Ancient World. Most of the time, the species exists at low densities in the solitarious form, characterized by relative cryptic and scattered individuals. At irregular intervals, it displays huge increases of population densities with actively aggregating and swarming individuals typical of the high-density gregarious form (Uvarov 1966). Eleven morphometrical taxa of *L. migratoria* with different propensities to display outbreaks were described, and considered as subspecies (Remaudière 1940a; Waloff 1940; COPR 1982). Comparisons of the levels of genetic diversity and differentiation in solitarious populations among historically outbreaking and non-outbreaking populations may provide important insight into the effect of outbreak events on population dynamics and structure of locusts. In particular, a populational homogenizing effect of outbreak events could be inferred if outbreaking populations were genetically less structured than non-outbreaking populations. Moreover, in populations recurrently fluctuating in their sizes, the long-term effective size,

which determines the overall amount of genetic drift, is expected to roughly correspond to the harmonic mean size over time and should be thus closer to the size during the remission period than that during the outbreak period (Motro and Thomson 1982). Consequently, if population sizes in solitarious populations were large/small, we would expect similarly high/low genetic diversities in historically outbreaking and non-outbreaking areas.

So far only a single study has described the genetic variation within and among solitarious populations, sampled four years after the last outbreak event, of a different locust species, the desert locust, *Schistocerca gregaria* (Ibrahim et al. 2000). This study, based on a single nuclear DNA marker, revealed large genetic diversity within populations (mean estimate of nucleotide diversity $\pi = 2.4\%$; Nei 1987) and substantial genetic variation among populations (mean estimate of net divergence $d_A = 0.3\%$; Nei 1987), despite the homogenizing potential of outbreak events. A number of hypotheses were presented to explain the latter pattern including (i) geographical isolation between breeding sites during outbreak and remission periods, (ii) failure of gregarious populations to establish at the end of the outbreak event, and (iii) the effect of repeated population extinction/recolonization events during the remission/outbreak periods (Ibrahim et al. 2000; 2001). However, a comparison of genetic variation of historically outbreaking populations with that of historically non-outbreaking populations has never been attempted in *S. gregaria* as well as in any other locust species, precluding clear inferences on the potential effect of outbreak events on genetic variation.

In this article, we compared the genetic variation within and among populations at 14 microsatellite markers for 24 *L. migratoria* solitarious populations located in three areas with contrasted historical patterns of outbreak events. Our extensive comparative microsatellite survey allowed useful inferences on the dynamics and structure of historically non-outbreaking and outbreaking populations of this pest species.

## Material & Methods

*Population sampling*

From autumns 2001 to 2004, we collected a total of 24 population samples (19 to 31 individuals per sample) distributed in three different areas: twelve populations in Madagascar (hereafter referred to as *MG* populations), four populations in Northern China (*NC*), and eight populations in Western Europe (*WE*) (Fig. 1). *MG*, *NC*, and *WE* sampled areas were relatively similar in size, with a maximal distance between two samples of 1533 km for Madagascar, 2503 km for Northern China, and 1046 km for Western Europe. Nine of the 24 population samples have been previously used for a worldwide genetic survey of *L. migratoria* by Chapuis et al. (in prep.; see Fig. 1).

The three studied areas are characterized by contrasted historical patterns of outbreak events. In *MG*, five widespread outbreaks and two upsurges controlled by insecticides, have been reported during the last century (Randriamanantsoa 1998). In *NC*, fourteen widespread outbreaks have been recorded from 1905 to 1959 in the Hebei region (Zhang and Li, 1999). Since this latter date, pest insecticides have contained outbreak events except in 1994 and 1998. Locust outbreaks have also been recurrently recorded in the XinJiang region (Chen 1991; Tanaka and Zhu 2005). The population samples from *MG*, which belong to the taxon *L. m. capito*, and from *NC*, which belong to the taxa *L. m. manilensis* and *L. m. migratoria*, were thus both considered as representative of frequently outbreaking populations. Malagasy and Chinese populations were sampled from one to six years after the last recorded outbreak event and most of them were composed of solitarious individuals. In *WE*, only one outbreak event, that probably originated from the Black Sea area in the XIV$^{th}$ century (Waloff 1940), and some rare, low intensity local outbreaks have been reported (region of Naples, Italy in 1936; Jannone 1947; region of Bordeaux, France in 1946-1948; Glize 1996). The *WE* samples, which encompass populations from the morphometrical taxa *L. m. cinerascens*, *L. m.*

*migratoria, L. m. gallica* and ‘Palavas’ form, were hence considered as representative of historically non-outbreaking populations.

Our sampling design is imbalanced, since we lack a second non-outbreaking population area that would serve as replicate for non-outbreaking populations. Unfortunately, historically non-outbreaking populations of *L. migratoria* are all located in Western Europe, with the exception of marginal populations located in small isolated islands (COPR 1982; Chapuis et al. in prep.). On the other hand, potential confounding ecological factors, such as number of generations, presence or absence of seasonal migration, or island vs. mainland area, were evenly distributed across the two considered patterns of outbreak events. The Malagasy intertropical environment is favourable to seasonal migration and a short generation time (four generations per year; Lecoq 1975) whereas the European and Chinese temperate zones are associated with facultative egg diapause and a larger generation time (one to two generations per year; Roubaud 1947 and Ma 1958).

*Genotyping*

Fourteen microsatellite loci (OZC9, OZC35, OZC76 in Zhang et al. 2003 and LM1-88, LM10-78, LM2-B, LM2-Λ, LM3-Ω, LMT-113, LMT-137, LMT-177, LM10-180, LM11-121, LMT-133 in Chapuis et al. 2005) were genotyped, using fluorescent PCR and a MegaBACE sequencing machine (Amersham Biosciences), as described in Chapuis et al. (in prep.). To minimize genotyping errors, a minimum of two replicate PCRs (Polymerase Chain Reactions) per sample per locus were conducted and alleles were scored, using the Genetic Profiler v.1.0 software (Amersham Biosciences), independently by two of the authors. Alleles included in the final consensus genotypes were observed at least twice; if observed only once, an additional PCR replicate was conducted. We included one negative control and eight positive

controls (samples with known genotypes) within each 96 PCR reaction set in order to check PCR contamination and to standardize genotypes among experiments.

*Null alleles*

Deviation from Hardy-Weinberg equilibrium (HWE) was tested for each locus separately and within each population using Fisher exact tests implemented in the software GENEPOP 3.3 (Raymond and Rousset 1995). Corrections for multiple tests were performed using the false discovery rate approach (Benjamini and Hochberg 1995). If observed genotype frequencies deviated significantly from HWE expectations, the program MICRO-CHECKER (van Oosterhout et al. 2004) was used to determine the most probable cause among nonpanmixia, genotyping error (e.g. short allele dominance, Wattier et al. 1998) and presence of null alleles. The presence of null alleles was inferred as the most probable cause if homozygote excess was homogeneously distributed across the homozygote classes. A maximum likelihood estimate of the frequency of null alleles (Dempster et al. 1977) was calculated using the program FreeNA (Chapuis and Estoup 2006).

*Upper levels of population structure and dispersal*

We applied a Bayesian analysis implemented in the program BAPS for estimating hidden substructure in the *MG*, *NC*, and *WE* areas separately (Corander et al. 2003; 2004). The method determines clusters of population samples minimizing Hardy-Weinberg and linkage disequilibrium within the clusters. In this method, the number of clusters is treated as an unknown parameter. This method has been shown very conservative in identifying population structure (Waples and Gaggiotti 2006). Computer simulations have shown that although BAPS better discriminate different gene pools in the presence of null alleles, its discrimination power remains low under high gene flow (Chapuis et al. in prep.). The method hence

determines the uppermost level of population structure, grouping together panmictic populations that exchange migrants at a high level. The data was analysed with both BAPS 2.0, which applies a Markov Chain Monte Carlo (MCMC) algorithm, and BAPS 3.1, which applies a stochastic optimisation, to infer population structuring.

We then assessed, separately for the *MG*, *NC*, and *WE* areas, whether (i) the different clusters of population samples provided by BAPS and (ii) the sampled populations within those clusters were the result of an ancestral population split with no gene flow (pure drift model) or of a balance between genetic drift and measurable gene flow under an infinite island model (mixed gene flow/drift model). Both models assume that the effects of microsatellite mutations are negligible. We used the 2MOD software (Ciofi et al. 1999) that applies a Markov chain Monte Carlo simulation approach with Metropolis-Hastings sampling to compare the likelihoods that each model (pure drift model vs. mixed gene flow/drift model) could have operated to produce the observed data. A set of three independent runs was carried out for each population set. For each run, 100,000 iterations were carried out and a burn-in period of 10,000 iterations was applied. The results presented are means of the three replicates. To our knowledge, the impact of the presence of null alleles in genotype datasets on the MOD2 approach has never been investigated. Unfortunately, this approach is computationally intensive and hence difficult to test using a large number of simulated datasets.

The BAPS and 2MOD treatments, in association with historical records of outbreak events, allowed identifying population sets homogeneous for both historical outbreak events and dispersal patterns. Such identification allowed more sensible comparisons of genetic variation among population sets.

*Comparison of genetic variation among population sets*

The mean expected heterozygosity( $\overline{H_E}$ , Nei 1987) of each population set and its confidence interval, calculated using 10,000 replicates of a bootstrap resampling procedure over loci, were computed using the program FSTAT (Goudet 1995). Computer simulations performed on five statistics summarizing genetic diversity within populations have shown that the presence of null alleles led to an underestimation of all such statistics, but that this bias was particularly low for the expected heterozygosity $H_E$ (Chapuis et al. in prep.).

For each population set, genotypic differentiation over all loci and populations was tested using Fisher exact tests (GENEPOP 3.3; Raymond and Rousset 1995). The level of differentiation between populations was quantified by computing Weir (1996) estimator of $F_{ST}$ across all populations of a same set. We used the so-called *ENA* correction method to efficiently correct for the positive bias induced by the presence of null alleles on $F_{ST}$ estimation (Chapuis and Estoup 2006). $F_{ST}^{\{ENA\}}$ values and their confidence intervals, calculated using 10,000 replicates of a bootstrap resampling procedure over loci, were computed using the package FreeNA (http://www.montpellier.inra.fr/CBGP/).

We used Wilcoxon signed-rank tests applied on the 14 single locus values of the statistics of interest ($H_E$ averaged over all populations and global $F_{ST}^{\{ENA\}}$ across all populations) to determine whether genetic diversity within populations and genetic differentiation between populations differed significantly between the population sets.

**Results**

*Null alleles*

A large excess of significant departures from HWE (i.e., heterozygote deficiency) were observed in 190 of the 336 single-locus exact tests. Specimens that failed to amplify at some loci did not yield PCR products after two or three attempts of PCR reactions whereas the

same DNA samples were successfully amplified at other loci. This strongly suggests the presence of null alleles for most loci and populations. In agreement with this, MICRO-CHECKER showed that the general excess of homozygotes is distributed across most allele size classes. Estimated frequencies of null alleles per locus per population ranged from 0 to 0.753, with frequencies averaged over populations and loci varying from 0.127 to 0.208 according to the population sets defined below (Table 1).

*Upper levels of population structure and dispersal*

The BAPS analysis provided the highest probability for a single population cluster in *MG* and in *NC* areas (Fig. 1). The pure drift model was rejected by the 2MOD method for both the twelve populations of the *MG* area and the four populations of the *NC* area, with strong and substantial support for gene flow/drift equilibrium for *MG* and *NC* populations, respectively (Table 2). The BAPS analysis provided the highest posterior probability for a substructure of five population clusters in the *WE* area (Fig. 1). The 2MOD method provided decisive evidence in favour of the pure drift model for the five population clusters inferred from BAPS in the *WE* area (Table 2). On the other hand, the 2MOD method strongly supported the mixed gene flow/drift model when applied to the four sampled populations belonging to *L .m. cinaerescens* and grouped together in a simple genetic cluster by BAPS (Table 1).

The BAPS and 2MOD analyses thus showed that the *WE* area includes two populational sets characterized by different patterns of dispersal: (i) a set of isolated populations distributed all over the sampled area (i.e., the four genetic clusters with a single population; hereafter referred to as *WE-I* set) and (ii) a set of populations distributed in Southern-Western Europe connected by gene flow (i.e., the cluster with the four sampled populations belonging to *L .m. cinaerescens*; *WE-C* set). We will hereafter compare genetic variation, within and among

populations, between five population sets: the historically outbreaking *MG* and *NC* population sets and the historically non-outbreaking *WE, WE-I* and *WE-C* population sets.

*Comparison of genetic variation among population sets*

All microsatellite loci genotyped in *L. migratoria* populations were highly variable, the mean expected heterozygosity varying from 0.82 to 0.87 (Table 3a). The two outbreaking population sets *MG* and *NC*, although located in regions with different ecological features, had similar levels of genetic variation within-population (Table 3b; $P > 0.058$). In the non-outbreaking *WE* area, the within-population genetic variation was significantly lower for populations of the set *WE-I* than for the set *WE-C* (Table 3b; $P \leq 0.013$). No significant differences were detected between all pairwise comparisons concerning population sets heterogeneous with respect to the outbreak pattern. Altogether these results indicated that genetic diversities were remarkably similar (and high) within outbreaking and non-outbreaking populations.

Global genetic differentiation between populations samples was highly significant within each population set ($P \leq 10^{-4}$) except for the *MG* population set ($P = 0.171$). In the *WE* area, level of genetic structure, as measured by the global $F_{ST}^{\{ENA\}}$ estimate, was about three fold greater for the *WE-I* populations than for the *WE-C* populations (Table 3a and 3b; $P = 0.001$). Both outbreaking population sets, although located in regions with different ecological features, had non significantly different low level of genetic structure (Table 3b; *MG* vs. *NC*: $P = 0.104$). Populations in areas with frequent and intense outbreak events were genetically less structured than populations in areas with uncommon and local outbreak events. Global $F_{ST}^{\{ENA\}}$ estimates differed significantly among all pairs of population sets heterogeneous with respect to the outbreak pattern ($P < 0.0001$; Table 3b). Global $F_{ST}^{\{ENA\}}$

estimates were four to thirty fold lower for the outbreaking population sets (*MG* and *NC*) than for the non-outbreaking population sets (*WE*, *WE-I*, and *WE-C*) (Table 3a).

## Discussion

*New insights into outbreak dynamics*

We found similar high genetic diversities within populations, as measured by $H_E$, in outbreaking and non-outbreaking populations of the migratory locust, suggesting similar harmonic means of effective population sizes. Genetic diversities within outbreaking populations were hence not strongly affected by recurrent increases of population sizes during outbreaking periods, as predicted by theory (Motro and Thomson 1982). The large enough amount of genetic variation observed within all populations indicates that solitarious locusts persist in populations of large effective size in both outbreaking and non-outbreaking areas.

The level of population differentiation, as measured by $F_{\mathrm{ST}}^{\{ENA\}}$, was an order of magnitude greater among Western European populations (*WE* set, including the *WE-I* and *WE-C* sets) than among Malagasy (*MG* set) and among Chinese populations (*NC* set). A lower level of genetic differentiation could reflect a higher gene flow among populations and/or a larger genetic diversity within populations (Slatkin 1995). However, we found estimates of genetic diversity to be similar among outbreaking and non-outbreaking populations, suggesting relatively similar harmonic means of effective population sizes. Our microsatellite survey hence provides strong evidence that gene flow among outbreaking populations is substantially greater than among non-outbreaking populations.

More intensive gene flow among populations of outbreaking areas may be the result of demographic and/or behavioural factors. First, as outbreaking populations experience recurrent demographic flushes, they will recurrently produce a larger number of effective migrants than non-outbreaking populations, for a given migration rate. Second, the high

densities of individuals during outbreak periods are likely to represent a strong demographic pressure for migration, which may result in an increase of the rate and/or the distance of effective migration in gregarious populations. In agreement with this, important changes in traits of migratory behaviour have been observed with the gregarious phase in locusts. For instance, it has been observed that gregarious adults, which fly at day, frequently overfly habitats that solitarious adults, which fly at night, can not reach because of low night temperatures (Uvarov 1966). Moreover, the actively aggregative forms have been interpreted as adaptations for long distance migration at least at the nymph stage (Ellis 1953).

The strong homogenizing effect of recent outbreak events in *L. migratoria* made unlikely, at least for this species, the hypotheses of a genetic isolation of solitarious and gregarious forms and of a drastic reduction of population viabilities at the time of outbreak declines (Ibrahim et al. 2000). Ibrahim et al. (2000) did not find evidence for a homogenizing effect of outbreak events on genetic variation in *S. gregaria*. *S. gregaria* and *L. migratoria* might differ in the populational characteristics of their gregarious and/or solitarious forms. Solitarious *S. gregaria* exists as small populations in patchy environments and are particularly prone to extinction because of climatic events (Ibrahim 2000) whereas sizes remain relatively large in solitarious *L. migratoria* populations (Lecoq 1975). However, the homogenising effect of outbreak events on genetic variation of *S. gregaria* may have been somewhat underrated due to the absence of comparison with genetic variation in non-outbreaking areas.

*Implications for preventive control*

Acting early in the outbreak process is essential to hamper upsurges with a limited amount of insecticide spraying, and hence to safeguard the environment (Lecoq 2001). Such preventive strategy is based on controlling first crowdings of solitarious locusts rather than vast and highly mobile bands or swarms of gregarious locusts. This strategy requires to determine the

origin of nascent outbreak populations and their subsequent movements, and the potential of genetic markers in doing so has been emphasized (e.g., see 2005-10 research priorities of the Australian Plague Locust Commission <http://www.affa.gov.au/aplc>). To what extent a genetic monitoring should be useful for survey and preventive management of locust pests depends on our statistical ability to detect migrants and infer their geographical origin. The traditional genetic approaches for inferring sources and routes of invasive processes, using *F*-statistics (e.g., Weir 1996) or assignment statistics (e.g., Rannala and Mountain 1997; Paetkau et al. 2004), might lack resolution when genetic differentiation between populations is low (e.g., Roeder et al. 2001), as found in historically outbreaking areas of *L. migratoria*. An efficient strategy that helps overcome the insufficient information in the genetic data is integrating in the analysis other type of data, as done by Gaggiotti et al. (2002; 2004) with demographic and geographical distance data. This method has been shown to be relatively insensitive to the level of genetic differentiation (Gaggiotti et al. 2004). Computer simulation studies are needed to assess the power of the latter method in the specific context of historically outbreaking areas of *L. migratoria*.

*Population structure in Western Europe*

We showed that the sampled Western European populations comprised a group of isolated populations and a group of populations connected by gene flow, that all belong to the subspecies *L. m. cinaerescens*. Isolated populations showed a lower genetic diversity and higher levels of population differentiation than connected populations. This could be explained by a smaller effective population size of isolated populations, which could be due to a lower carrying capacity of their habitat. Northern and Western Europe, correspond to limits of the species range, and is hence likely to worse bit the ecological optimum of the species than the Mediterranean coast, due to colder climatic conditions. For instance, the sampled

French population of *L. m. migratoria*, is the Northerner reported population in Western Europe (P. Meunier, pers. comm.). Conversely, a lower carrying capacity of habitats is unlikely to explain lower genetic diversity for the populations belonging to the Palavas form. Indeed, this form is found in a small area of the Hérault region that comprises lagoon ecosystems that offer wet habitats particularly favourable to the migratory locust (Remaudière 1940a).

The 2MOD method suggests that the lower genetic diversity and higher levels of population differentiation of isolated populations of Western-Europe could also be explained by an absence of gene flow. This absence of gene flow could be due to geographic isolation for the two Northern populations but not for the populations of the Palavas form since the latter are found on the Mediterranean coast along which populations of the subspecies *L. m. cinerascens* are commonly found (Fig. 1). The pre and/or post-mating isolation mechanism(s) involved in the apparent absence of effective gene flow of the Palavas form with neighbouring *L. m. cinaerescens* populations are unknown. Adults of the Palavas form are notorious for their remarkably larger size than that of their neighbours (Remaudière 1940b; personal observation). It is conceivable that premating isolation through size-assortative mating have evolved, rendering migration between the Palavas form and *L. m. cinerascens* ineffective. Assortative mating by size is common in insects (reviewed in del Castillo 1999), and can be generated by mechanisms such as mate choice, mate availability, and mating constraints. Field and laboratory experiments on mating behaviour are needed to test for this size-assortative mating hypothesis.

**Acknowledgments**

We are grateful to L. Baoping, P. Fontana, A. Foucart, D. Hunter, P. Meunier, D. Llucià Pomares, G. Reynaud, X. Shengquan, F. Tami, S. Wangpeng, D.-X. Zhang, and G. Zhiyong

for providing us with *L. migratoria* population samples. M.P.C. was supported by a grant from the CNRS. This work was partly funded by INRA (scientific SPE department) and by CIRAD. YM acknowledges support from CNRS and IRD.

**Literature cited**

Bailey NW, Gwynne DT, Ritchie MG (2005) Are solitary and gregarious Mormon crickets (*Anabrus simplex*, Orthoptera, Tettigoniidae) genetically distinct? *Heredity* **95**, 166-173.

Benjamin Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* **57**, 289-300.

Berthier K, Charbonnel N, Galan M, Chaval Y, Cosson JF (2006) Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. *Molecular Ecology* **in press**.

Chapman RF (1976) A biology of locusts. *Studies in Biology* **71**.

Chapuis M-P, Estoup A (2006) Microsatellite null alleles and estimation of population differentiation. *Molecular and Biological Evolution* **in press**.

Chapuis M-P, Lecoq M, Loiseau A, Piry S, Michalakis Y, Estoup A (in prep.) Worldwide microsatellite genetic variation in a pest species prone to null alleles, *Locusta migratoria*: congruence with current taxonomy and outbreak patterns.

Chapuis M-P, Loiseau A, Michalakis Y, Lecoq M, Estoup A (2005) Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria*. *Molecular Ecology Notes* **5**, 554-557.

Chen Y-L (1991) The migratory locust, *Locusta migratoria*, and its asiatic subspecies. *Ortopterists' Society & Lyman Entomological Museum, Québec*.

Ciofi C, Beaumont MA, Swingland IR, Bruford MW (1999) Genetic divergence and units for conservation in the Komodo dragon *Varanus komodensis*. *Proceedings of Royal Society of London series B* **266**, 2269-2274.

COPR (1982) The locust and grasshopper agricultural manual. *Centre for Overseas Pest Research, London.*

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**, 2363-2369.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

Del Castillo RC, Nunez-Farfan J, Cano-Santana Z (1999) The role of body size in mating success of *Sphenarium purpurascens* in Central Mexico. *Ecological Entomology* **24**, 146-155.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1-38.

Duranton JF, Launois M, Launois-Luong MH, Lecoq M (1979) Les voies privilégiées de déplacement du Criquet migrateur malgache en phase solitaire. *Bulletin d'Ecologie* **10**, 107-123.

Ehrich D, Jorde PE (2005) High genetic variability despite high-amplitude population cycles in lemmings. *Journal of Mammalogy* **86**, 380-385.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among population of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical Applied Genetics* **92**, 832-839.

Ellis PE (1953) The gregarious behaviour of marching *Locusta migratoria migratorioides* (R. & F.) hoppers. *The journal of experimental biology* **30(2)**, 214-234.

Gaggiotti OE, Brooks SP, Amos W, Harwood J (2004) Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. *Molecular Ecology* **13**, 811-825.

Gaggiotti OE, Jones F, Lee WM, *et al.* (2002) Patterns of colonization in a metapopulation of grey seals. *Nature* **416**, 424-427.

Glize E (1996) L'invasion des criquets dans les Landes. *Bulletin de la Société de Borda* **121**, 173-194.

Hunter DM (2004) Advances in the control of locusts (Orthoptera: Acrididae) in eastern Australia: from crop protection to preventive control. *Australian Journal of Entomology* **43**, 293-303.

Ibrahim KM (2001) Plague dynamics and population genetics of the desert locust: can turnover during recession maintain population genetic structure? *Molecular Ecology* **10**, 581-591.

Ibrahim KM, Sourrouille P, Hewitt GM (2000) Are recession populations of the desert locust (*Schistocerca gregaria*) remnants of past swarms? *Molecular Ecology* **9**, 783-791.

Jannone G (1947) Italy: numerous outbreaks of *Locusta migratoria* L. phase gregaria (typical) and phases of transition in the province of Naples. *International Bulletin of Plant Protection* **10**, 218-219.

Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166**, 1081-1092.

Lecoq M (1975) Les déplacements par vol du criquet migrateur malgache en phase solitaire : leur importance sur la dynamique des populations et la grégarisation. *Thèse de Doctorat d'Etat ès Sciences (Université Paris XI, Orsay). Ministère de la coopération (Paris) : 272 p.*

Lecoq M (2001) Recent progress in Desert and Migratory Locust management in Africa. Are preventive actions possible? *Journal of Orthoptera Research* **10**, 277-291.

Ma S-C (1958) The population dynamics of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China. *Acta Entomologica Sinica* **8**, 1-40.

Motro U, Thomson G (1982) On heterozygosity and the effective size of populations subject to size changes. *Evolution* **36**, 1059-1066.

Nei M (1987) Molecular Evolutionary Genetics. *(Columbia University Press)*.

Paetkau A, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Randriatmanantsoa M (1998) Manuel sur la lutte antiacridienne. *Direction de la protection des végétaux et Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, Antananarivo*.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* **94**, 9197-9221.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Heredity* **86**, 248-249.

Remaudière G (1940a) Contribution à l'étude des *Locusta migratoria migratoria* L. ph. solitaria de la région de Palavas (Hérault) 2ère partie. *Revue de Pathologie Végétale et d'Entomologie Agricole de France* **17**, 220-235.

Remaudière G (1940b) Contribution à l'étude des *Locusta migratoria migratoria* L. ph. solitaria de la région de Palavas (Hérault) 1ère partie. *Revue de Pathologie Végétale et d'Entomologie Agricole de France* **17**, 147-163.

Roeder AM, Marshall RK, Mitchelson AJ, *et al.* (2001) Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. *Molecular Ecology* **10**, 1645-1656.

Roubaud E (1947) Au sujet du problème acridien en France, différences radicales évolutives entre les peuplements du Criquet migrateur (*Locusta migratoria* L.) des Landes du

Sud-Ouest et ceux de la région méditerranéenne. *Comptes Rendus des Séances de l'Académie des Sciences* **225**, 909-911.

Rousset F (2003) Effective size in simple metapopulation models. *Heredity* **91**, 107-111.

Servedio MR (2004) The evolution of premating isolation: local adaptation and natural and sexual selection against hybrids. *Evolution* **58**, 913-924.

Simpson SJ, Sword GA, Lecoq M (submitted) Nipping locust outbreaks in the bud: the importance of targeted, local scale monitoring. *Journal of Orthoptera Research*.

Slatkin M (1994) Linkage disequilibrium in growing and stable-populations. *Genetics* **137**, 331-336.

Tanaka S, Zhu D-H (2005) Outbreaks of the migratory locust *Locusta migratoria* (Orthoptera: Acrididae) and control in China. *Applied Entomological Zoology* **40**, 257-263.

Thomas MB (1999) Ecological approaches and the development of "truly integrated" pest management. *Proceedings of the National Academy of Sciences USA* **96**, 5944-5951.

Uvarov BP (1966) Grass-hoppers and Locusts. *Cambridge University Press* **Volume 1**.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Waloff ZV (1940) The distribution and migrations of *Locusta* in europe. *Bulletin of Entomological Research* **31**, 211-246.

Waples RS, Gaggiotti OE (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **in press**.

Wassenaar LI, Hobson KA (1998) Natal origins of migratory monarch butterflies at wintering colonies in Mexico: New isotopic evidence. *Proceedings of the National Academy of Sciences USA* **95**, 15436-15439.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998) Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology* **7**, 1569-1573.

Weir BS (1996) Genetic Data Analysis II. *Sinauer: Sunderland, MA.*

Whitlock M (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* **46**, 608-615.

Zhang DX, Yan LN, Ji YJ, *et al.* (2003) Isolation, characterization and cross-species amplification of eight microsatellite DNA loci in the migratory locust (*Locusta migratoria*). *Molecular Ecology Notes* **3**, 483-486.

Zhang Z, Li D (1999) A possible relationship between outbreaks of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China and the El Nino episodes. *Ecological Research* **14**, 267-270.

## Tables

**TABLE 1. Heterozygosities, HWE and null allele frequency estimation.** $n$, number of populations; $\overline{H_O}$, mean observed heterozygosities; $\overline{H_E}$, mean expected heterozygosities (Nei 1987); $D_{HWE}$, mean number of loci showing significant deviation (i.e., heterozygote deficiency) from Hardy-Weinberg equilibrium (HWE) after false discovery rate correction; $\overline{\hat{r}_D}$, mean null allele frequencies computed with the method of Dempster et al. (1977) on the original genotype datasets. All means were computed across loci and populations. Note that estimating gene frequencies using the maximum likelihood algorithm of Dempster et al. (1977) decreased the number of significant departures from *HWE*, after false discovery rate correction, from 190 to 0. *MG*, Madagascar; *NC*, Northern China; *WE*, Western Europe; *WE-I*, group of isolated populations in Western Europe; *WE-C*, group of connected populations in Western Europe. See text for details.

| Population set | $n$ | $\overline{H_O}$ | $\overline{H_E}$ | $D_{HWE}$ | $\overline{\hat{r}_D}$ |
|---|---|---|---|---|---|
| *MG* | 12 | 0.649 | 0.833 | 8.1 | 0.127 |
| *NC* | 4 | 0.570 | 0.834 | 9.5 | 0.208 |
| *WE* | 8 | 0.581 | 0.846 | 9.5 | 0.199 |
| *WE-I* | 4 | 0.569 | 0.818 | 9.5 | 0.191 |
| *WE-C* | 4 | 0.594 | 0.873 | 9.5 | 0.206 |

**TABLE 2. Evidence in favour of models of Pure drift or Mixed gene flow/drift within and among BAPS genetic clusters.** Genetic clusters of population samples were assessed without a priori information by the Bayesian method of Corander et al. (2003), separately for the *MG*, *NC*, and *WE* areas (see Fig. 1). The two relative likelihoods of the models of Pure drift and Mixed gene flow/drift were computed as the proportions of data points representative of these models (Ciofi et al. 1999). Bayes factor values are the ratios of the largest relative likelihood on the smallest relative likelihood. The model with the largest relative likehood is "substantially" (*), "strongly" (**) or "decisively" (***) supported when Bayes factor values range from 3.2 to 10, from 10 to 100, and above 100, respectively (Kass and Raferty 1995). *MG*, Madagascar; *NC*, Northern China; *WE*, Western Europe .

| | | Relative likelihood | | Bayes factor |
|---|---|---|---|---|
| | | Pure drift model | Mixed gene flow/drift model | |
| *MG* | Within genetic cluster (i.e., among all 12 sampled populations) | 0.22 | 0.78 | 3.5* |
| *NC* | Within genetic cluster (i.e., among all 4 sampled populations) | 0.02 | 0.98 | 49.0** |
| *WE* | Among the 5 genetic clusters | 1.00 | 0.00 | ∞*** |
| | Within genetic cluster (i.e., among the 4 *L.m.cinaerescens* populations) | 0.12 | 0.88 | 7.3* |

**TABLE 3. Comparison of genetic variation among population sets.**

$\overline{H_E}$, expected heterozygosities averaged over loci and populations (Nei 1987); $F_{ST}^{\{ENA\}}$, Weir's (1996) $F_{ST}$ computed across loci and populations taking into account the presence of null alleles (Chapuis and Estoup 2006); *WE*, Western Europe; *WE-I*, group of isolated populations in Western Europe; *WE-C*, group of connected populations in Western Europe; *MG*, Madagascar; *NC*, Northern China. (a) Estimated values of $\overline{H_E}$ and $F_{ST}^{\{ENA\}}$ for the different population sets. 95% confidence intervals computed by bootstrap resampling over loci are given between square brackets. (b) Results of Wilcoxon signed rank tests (*P*-values). Significant deviations from the null expectation of no difference between population sets ($P < 0.05$) are shown in bold characters.

**(a)**

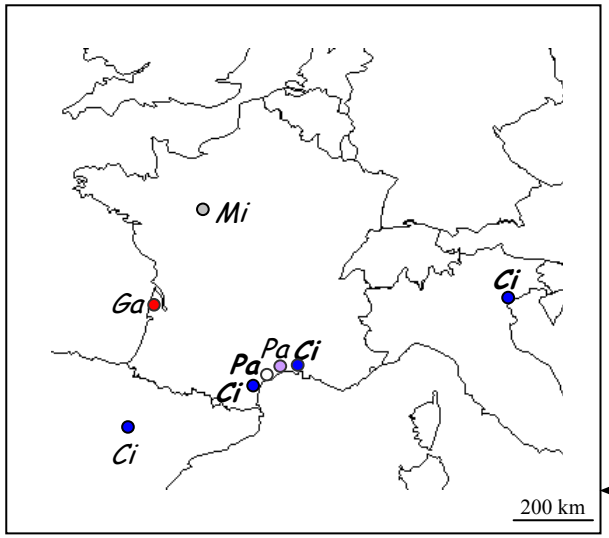
| | Outbreaking population sets | | Non-outbreaking population sets | | |
|---|---|---|---|---|---|
| | *MG* | *NC* | *WE* | *WE-I* | *WE-C* |
| $\overline{H_E}$ | 0.83 [0.77-0.91] | 0.83 [0.74-0.91] | 0.85 [0.81-0.89] | 0.82 [0.78-0.85] | 0.87 [0.82-0.91] |
| $F_{ST}^{\{ENA\}}$ | 0.002 [0.000-0.004] | 0.005 [0.002-0.026] | 0.039 [0.035-0.046] | 0.058 [0.047-0.071] | 0.021 [0.015-0.027] |

**(b)**

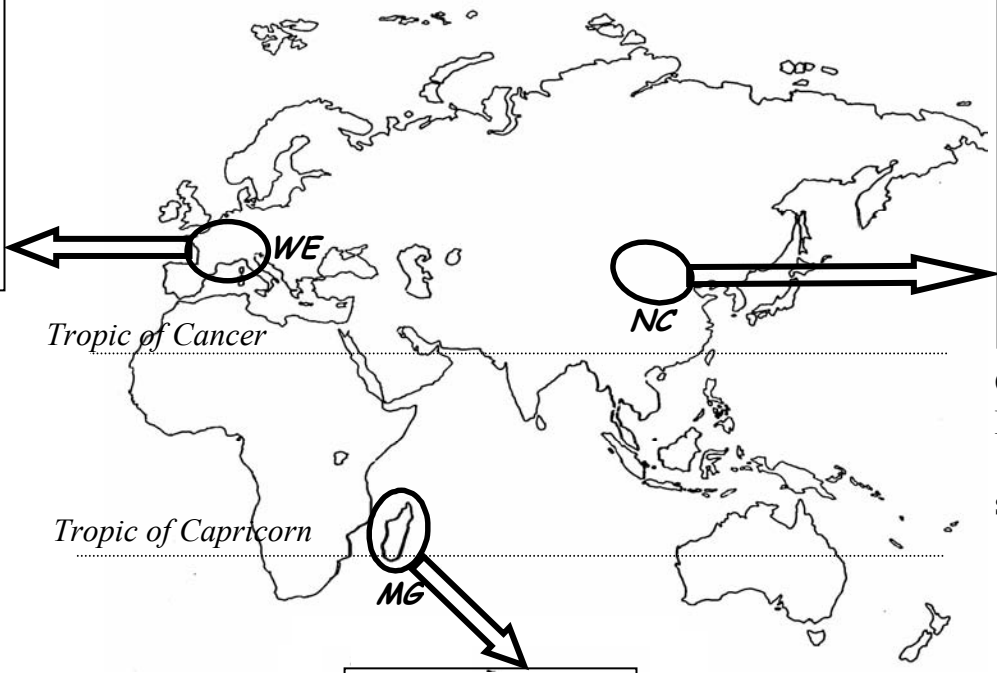
| | Within *WE* | Pair of outbreaking population sets | Pairs of population sets with heterogeneous outbreak patterns | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *WE-I* vs. *WE-C* | *MG* vs. *NC* | *WE* vs. *MG* | *WE* vs. *NC* | *WE-I* vs. *MG* | *WE-I* vs. *NC* | *WE-C* vs. *MG* | *WE-C* vs. *NC* |
| $\overline{H_E}$ | **0.003** | 0.553 | 0.761 | 0.594 | 0.426 | 0.119 | 0.296 | 0.078 |
| $F_{ST}^{\{ENA\}}$ | **0.001** | 0.104 | **<0.001** | **<0.001** | **<0.001** | **<0.001** | **<0.001** | **<0.001** |

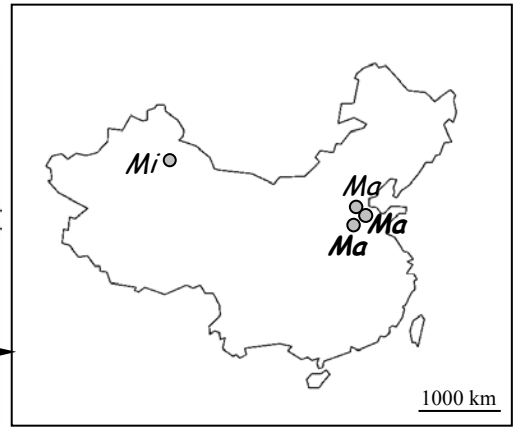
**Figure legends**

**FIG. 1. Geographical origins, taxonomical units, and genetic clustering of sampled populations of the migratory locust.** *WE*, Western Europe; *MG*, Madagascar; *NC*, Northern China; *Ca*, *L. m. capito*; *Ci*, *L. m. cinaerescens*; *Ga*, *L. m. gallica*; *Ma*, *L. m. manilensis*; *Mi*, *L. m. migratoria*; *Pa*, 'Palavas' form. Sampled sites with similar colours belong to the same genetic cluster as assessed by the Bayesian clustering method of Corander et al. (2003). The samples not previously used in the worldwide genetic survey of *L. migratoria* by Chapuis et al. (in prep.) are in bold.

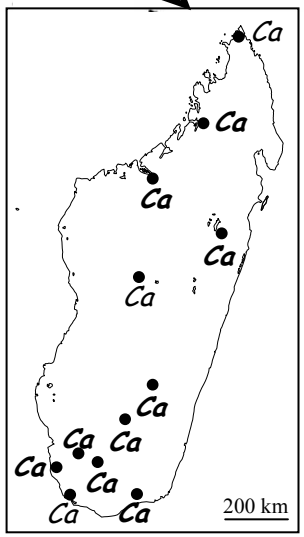



**Outbreak pattern**: uncommon and local
**Ecological characteristics**:
1-2 generations per year, absence of seasonal migration




**Outbreak pattern**: frequent and intense
**Ecological characteristics**:
1-2 generations per year, absence of seasonal migration




**Outbreak pattern**: frequent and intense
**Ecological characteristics**: 4 generations per year, presence of a seasonal migration

**FIG. 1.**

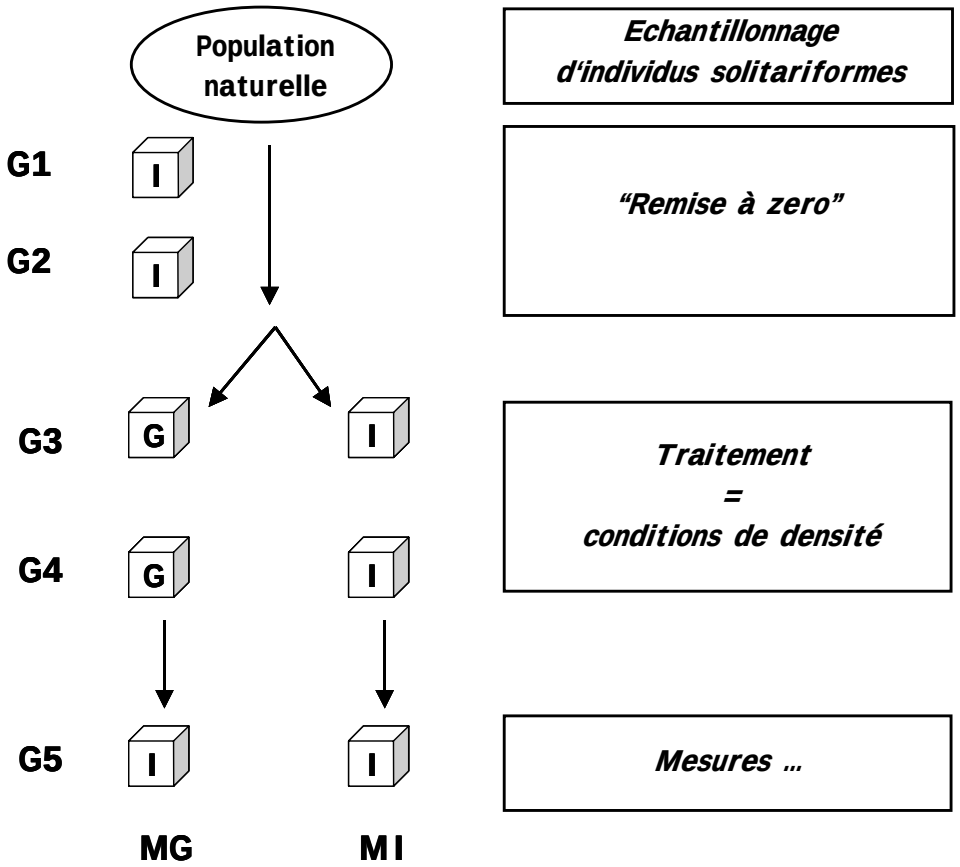
# **Chapitre III.** Evolution expérimentale de la grégarisation

L'étude de l'évolution expérimentale de la grégarisation a été réalisée chez *L. migratoria* dans le but d'évaluer (i) la possibilité d'un déterminisme génétique de la propension à grégariser et (ii) l'effet de la grégarisation sur les traits d'histoire de vie. Quelques rares études ont précédemment tenté d'évaluer chez les locustes la variation génétique du changement phasaire (*e.g.*, Heifetz *et al.* 1994 ; Sword 2003) ainsi que la variation de traits d'histoire de vie avec la phase (*e.g.* Norris 1950). Cependant, leurs résultats sont souvent imprécis ou conflictuels. Une des raisons tient au protocole choisi d'induction de la phase en élevage. Ce chapitre expose donc les approches méthodologiques utilisées d'une part dans les études précédentes et d'autre part au cours de cette thèse, avant de présenter une synthèse des analyses et résultats des deux questions traitées. Ces deux questions sont respectivement détaillées dans les manuscrits 5 et 6 inclus à la fin du chapitre.

## 1. APPROCHE METHODOLOGIQUE

Le changement de phase chez les locustes est classiquement considéré comme l'expression de la plasticité phénotypique, *i.e.* la capacité des individus à produire des phénotypes différents, du solitaire au grégaire, quand les conditions environnementales changent de l'isolement au groupement. En effet, le changement de phase a généralement été analysé par la comparaison d'échantillons élevés en isolement ou en groupement, que le trait étudié soit écologique, moléculaire, physiologique, comportemental ou morphologique. Cependant, le changement de phase est distinct de la plasticité phénotypique *sensu stricto*, puisque le phénotype et l'environnement maternels ont une empreinte dans la variation phénotypique de la descendance (pour plus de détail voir la section 2.3.2 de l'introduction). En conséquence, l'approche classiquement utilisée pour mesurer le changement de la phase solitaire à la phase grégaire confond les effets de densité proximaux, qui résultent de l'expérience propre d'un individu, avec les effets environnementaux et génétiques transmis. Dans cette thèse, nous avons réalisé une expérience d'élevage en laboratoire visant à distinguer les effets de l'histoire des conditions de densité des générations parentales de ceux des conditions de densité contemporaines aux individus mesurés. Le protocole choisi dans ce but est présenté dans la Fig. 12.

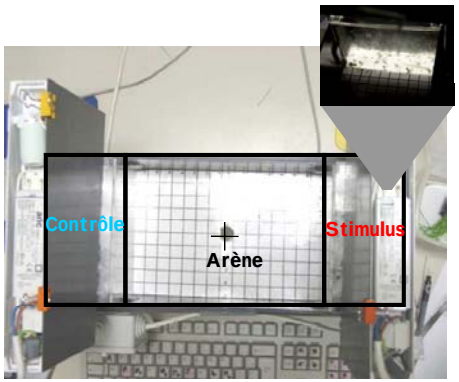
**Fig 12. Protocole expérimental de mesure du changement de phase.** *I* , conditions d'isolement ; *G*, conditions de groupement ; *G1* à *G5*, 1ère à 5ème génération de l'élevage en laboratoire.






Photographies d'individus élevés en cage de groupe (G) et en isolement (I) mettant en évidence les différences colorimétrique et morphométrique (e.g., voir la forme de la carène) entre les deux formes.

***…du comportement, de la morphologie, et de traits d'histoire de vie.***




Photographie de la cage utilisée pour les mesures comportementales. Le stimulus est constitué d'environ 80 larves de la même population. L'individu est introduit au centre de l'arène à l'aide d'une seringue.

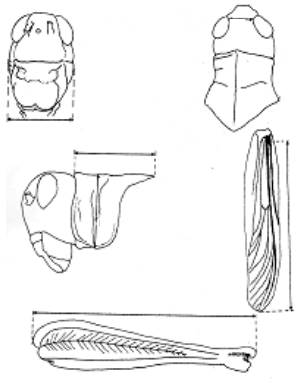


Illustration des mesures réalisées pour calculer les variables morphométriques.

Exemple de scan de larves à l'émergence réalisé pour mesurer, à l'aide d'un logiciel d'analyse d'images, la longueur de fémur de la descendance des femelles.

## 2. VARIATION GENETIQUE DE LA PROPENSION A GREGARISER

Etant donnée l'importance du processus de grégarisation dans la formation et le maintien des évènements de pullulations, ses déterminants ont été largement étudiés. Alors que les paramètres environnementaux ont rapidement émergé comme des facteurs clés de la grégarisation, l'implication d'un déterminisme génétique est plus confuse. Nous avons caractérisé la réponse des traits phasaires comportementaux et morphométriques à une histoire de groupement ou d'isolement de deux populations de *L. migratoria* avec des patrons historiques de pullulations contrastés (Manuscrit 5). Nous avons trouvé qu'une population malgache, historiquement pullulante, exprimait un niveau de grégarisation plus élevé que celui d'une population française, historiquement non-pullulante. Comme nous avons élevé les deux populations dans un environnement commun (isolement), la variation de la propension à grégariser résulte probablement d'un processus génétique adaptatif. La forte propension à grégariser à Madagascar peut être due à une pression de sélection pour maintenir la plasticité phénotypique. Le front inter-tropical des pluies à Madagascar est en effet responsable d'une grande variation saisonnière des densités des populations de locustes. Il est probable que la propension à grégariser ait été au moins en partie perdue dans la population française par dérive génétique depuis que la sélection pour le caractère grégaire a été relâchée, et/ou par contre-sélection si la plasticité phénotypique est coûteuse.

## 3. ROLE DE LA PHASE DANS LE CONTROLE DES PULLULATIONS

Nous avons voulu savoir si les forts taux de croissance des populations grégaires en période de pullulation pouvaient être expliqués par de meilleurs taux de survie et/ou de reproduction de la phase grégaire relativement à la phase solitaire (Manuscrit 6). Pour cela, nous avons évalué l'influence des phases solitaire et grégaire sur la survie et la fécondité des mères, ainsi que sur la qualité et le sexe-ratio de leurs descendants, chez une population malgache avec une histoire de pullulations fréquentes. Nous avons mesuré, en conditions homogènes d'isolement contrôlées en laboratoire, un groupe d'un échantillon de cette population avec une histoire d'isolement et un groupe du même échantillon avec une histoire de groupement (Fig. 12). Nous avons montré que la

population malgache augmentait son potentiel reproducteur (reproduction plus précoce et meilleur qualité de la descendance) lorsque les générations parentales étaient élevées en condition de groupement. Comme le groupement des générations parentales a conduit à une grégarisation morphométrique et comportementale significative, nous pouvons conclure que les individus de la phase grégaire ont un meilleur potentiel reproducteur que les individus de la phase solitaire chez *L. migratoria*. En revanche, nous n'avons pas mis en évidence d'effets de la phase héritée maternellement sur le développement ou la survie des femelles, ni sur le sexe-ratio de leur descendance.

Il faut noter que, pour des raisons logistiques, nous avons choisi de mesurer la survie et la fécondité des différentes phases uniquement en conditions d'isolement. Ces conditions d'isolement représentent des conditions de disponibilité en ressources optimales et des conditions de stress minimes, pour lesquelles les compromis entre traits d'histoire de vie sont plus difficilement détectables (Zera et Harshman 2001). D'autre part, mesurer les traits en conditions d'isolement est susceptible de contre-balancer l'expérience maternelle des conditions de groupement, minimisant ainsi les différences entre les groupes d'histoire en conditions de groupement et d'isolement. Ces propriétés rendent notre approche méthodologique très conservative au point de « masquer » certains effets sur les traits d'histoire de vie.

L'ensemble des résultats suggèrent que le processus de grégarisation peut avoir un rôle dans la croissance numérique des populations durant le développement et/ou le maintien des pullulations chez *L. migratoria*. Cependant, il faut noter que les effets maternels observés dans notre expérience sont faibles. De plus, les paramètres de croissance des populations, montrent souvent une variation induite par les conditions environnementales de l'individu lui-même. Il est malheureusement difficile en termes logistiques de mesurer les traits d'histoire de vie d'individus élevés en conditions de groupement, qui reflètent en élevage les conditions de fortes densités des populations grégaires, car l'unité de mesure n'est plus une cage individuelle mais une cage de plus grande taille contenant un nombre important d'individus.

## MANUSCRIT 5 – GENETIC VARIATION FOR PARENTAL EFFECTS ON PROPENSITY TO GREGARISE IN *LOCUSTA MIGRATORIA*

Marie-Pierre Chapuis, Arnaud Estoup, Arnaud Augé-Sabatier, Antoine Foucart, Michel Lecoq et Yannis Michalakis

Soumis à BMC Evolutionary Biology.

# Genetic variation for parental effects on propensity to gregarise in *Locusta migratoria*

Marie-Pierre Chapuis[1, 2, 3§], Arnaud Estoup[1], Arnaud Augé-Sabatier[1], Antoine Foucart[3], Michel Lecoq[3] and Yannis Michalakis[2]

[1]Centre de Biologie et de Gestion des Populations, Institut National de la Recherche Agronomique, Campus International de Baillarguet CS 30016, 34988 Montferrier/Lez, France

[2]Génétique et Evolution des Maladies Infectieuses, UMR 2724 CNRS-IRD, IRD, 911 avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France

[3]Centre de coopération internationale en recherche agronomique pour le développement, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

[§]Corresponding author

Email addresses:

MPC: chapuimp@ensam.inra.fr

AE: estoup@ensam.inra.fr

AAS: arnaudauges@yahoo.fr

AF: antoine.foucart@cirad.fr

ML: lecoq@cirad.fr

YM: Yannis.Michalakis@mpl.ird.fr

# Abstract

## Background

Environmental parental effects can have important ecological and evolutionary consequences, yet little is known about genetic variation in the plastic response of offspring phenotype to parental environment among populations. This type of variation may lead to rapid population phenotypic divergence and facilitate speciation. With respect to density-dependent phenotypic plasticity, some Acridid species, called locusts, present a spectacular developmental and behavioural shift, called phase change. Given the significance of phase change in locust outbreaks and control, its triggering processes have been widely investigated. Whereas crowding both within the offspring and parent lifetimes has rapidly emerged as a primary driving factor of phase change, less is known about a potential intra-specific genetic variation in the expression of phase change. We therefore carried out a laboratory experiment which enabled us to characterize the response of phase-related traits to laboratory controlled parental density for two natural genetically distinct populations of *Locusta migratoria* with contrasted historical patterns of outbreak events.

## Results

We found evidence that the historically outbreaking population of *L. migratoria* expressed larger parentally-inherited phase change than the historically non-outbreaking population.

## Conclusions

Because both populations were raised in a common environment during our experiment, an adaptive genetically-based process must be responsible for this

variation in the propensity to gregarise. This result also suggests a role for density-dependent parental effects in driving evolution of phase change within the family Acrididae.

## Background

Phenotypic plasticity, the capacity of a genotype to produce different phenotypes in different environments, is ubiquitous in nature [1]. It is now clear in many cases that phenotypic plasticity is adaptive, allowing organisms to exploit temporally and spatially heterogeneous environments [2]. When there are geographic differences in temporal variability in the environmental conditions, there is a potential for the evolution of population differentiation in phenotypic plasticity [3-5].

Non-genetic parental effects deserve a particular attention in light of current theoretical interest in the evolution of plasticity [6], since they can be viewed as the plastic response of the offspring phenotype to the parental environment [7]. Although parental effects on reaction norms for morphological, behavioural, and physiological traits have been investigated in a variety of organisms [reviewed in 8 for insects], only recently have environmental parental effects been viewed as a form of plasticity that may in itself evolve [7, 9]. If environmental parental effects are important in natural populations, genetic differentiation among populations in reaction norms induced by parental environments may be fairly common, as it is for other types of phenotypic plasticity [10]. Unfortunately, little is known about the extent of such genetic variation among populations occupying habitats with different levels of selection. This is especially true for animals (for plants see [7, 10]).

Many insect species exhibit plastic changes in physiology, behaviour and morphology in response to crowding [11]. Crowding-dependent phenotypic plasticity

is expressed to varying degrees among members of the Coleoptera, Lepidoptera, Hemiptera, Homoptera, and Orthoptera [11]. However, the ability to gregarise where individuals undergo extreme phase change from the solitarious, at low population density, to the gregarious, at high population density, is a defining feature of locusts [12-14]. This remarkable phase change has been interpreted as an adaptation for migration [15] or high population density, that mediates high risks of predation [16], disease [17], or competition for food and female resources [18, 19]. Environmental parameters have rapidly emerged as primary driving factors of phase change in locusts [20]. Phase change depends primarily on population density within both the offspring and parent lifetimes [13, 21, 22]. The change induced by the crowding environment of the parent drives a positive loop towards further and further gregarisation across generations; it is thus central to the extreme nature of the density-dependent phase change in locusts. What remains unclear, however, is the potential implication of a genotype-dependent variation in the expression of phase change, and in particular in response to the parental environment.

The possibility of a genetic determinism on phase trait variation deserves to be raised especially in *L. migratoria*, which presents geographic variation in its propensity to display population outbreaks [23]. Among *L. migratoria* populations, swarming and the expression of gregarisation are highly correlated [24]. Population heterogeneity in the propensity to gregarise reflects environmental heterogeneity in the temporal variation in locust population densities. Most of the time, the species exists at low densities in the solitarious phase. In historically outbreaking areas, the species displays at irregular intervals huge increases of population densities with swarming individuals typical of the gregarious phase. It has been shown that historically non-outbreaking populations of *L. migratoria* also change phase-related

traits in response to laboratory controlled conditions of crowding [25, 26]. However, under the assumption that propensity to phase change is a genetically-based adaptive response to crowding, we predicted that a historically outbreaking population would more pronouncedly change its phase across generations in response to parental crowding than a historically non-outbreaking population.

To test this prediction, we carried out an experiment investigating the norms of reaction of two commonly used phase-related sets of traits, morphometry of adults and behaviour of nymphs, to parental density conditions for two natural closely-related populations of *L. migratoria* with contrasted historical patterns of outbreak events, namely a historically frequently outbreaking population from Madagascar and a historically non-outbreaking population from France. We first reared both populations under isolated conditions for two generations to control for natural parental histories, then reared both populations under isolated or crowded conditions for two generations, and finally observed progeny phenotypes for phase traits under homogeneous conditions (i.e., isolation). Because all assayed individuals are raised in the same environment, our design eliminates proximal environment effects and allows us to ascribe potential differences to population differences. We found that the historically outbreaking population of *L. migratoria* expressed larger parentally-inherited phase change than the historically non-outbreaking population, providing new insight into the evolution of phase change.

## Results

Before proceeding to the analysis of the interaction Populational origin x Rearing history specific effect, we first briefly report the "baseline" differences in

morphometry and behaviour between the French and Malagasy populations (the Populational origin effects in the full factorial MANOVAs with Box-Cox transformed data; $P < 0.0001$). French locusts had eyes more spaced, a smaller ratio of the length of the fore wing to the length of the hind femur, and a larger ratio of the length of the hind femur to the maximum width of the head than Malagasy locusts (first canonical function *CF1* of the canonical discriminant analysis of the morphometric data). French locusts were also more repelled by other locusts, had a more tortuous track, spent more time walking, spent less time climbing and swaying, and jumped more frequently than Malagasy locusts (*CF1* of the canonical discriminant analysis of the behavioural data). Such inter-population differences in the solitarious phenotypes of *L. migratoria* have already been reported by previous studies [26] and are expected through local adaptation to the French temperate vs. Malagasy inter-tropical habitats.

The Malagasy and French populations significantly differed in the variation of morphometric and behavioural phase traits with Rearing history (the interactions Populational origin x Rearing history in the full factorial MANOVAs with Box-Cox transformed data; $P = 0.0008$ and $P < 0.0001$, respectively). Because canonical discriminant analysis is identical to multivariate analysis of variance while providing the relationships between groups, we discuss the canonical functions rather than the details of the MANOVA tests (details of this analysis are provided in Additional file 1).

The first two canonical functions, obtained by the canonical discriminant analysis of the measurements of the four morphometric variables, yielded significant differences among group centroids (*CF1*: Wilks' Lambda; $\chi^2_{12} = 95.6; P < 0.0001$; second canonical function *CF2*: Wilks' Lambda; $\chi^2_6 = 32.3; P < 0.0001$) and accounted for 99 % of the variation between groups (Figure 1a). Rearing history did

not significantly affect morphometric traits of the French population, contrary to the Malagasy population (test on Mahalanobis distances: $P = 0.0825$ and $P < 0.0001$ respectively). Malagasy *F/C*, *DE*, and *PH/PL* values decreased with crowding history (forward stepwise selection procedure with $F = 2.706$ on Malagasy groups with histories of isolation and crowding), in agreement with previous morphometrical studies [26].

The first two canonical functions, obtained by canonical discriminant analysis of the measurements of the eleven behavioural variables, yielded significant differences among group centroids (*CF1*: Wilks' Lambda; $\chi^2_{33} = 208.9; P < 0.0001;$ *CF2*: Wilks' Lambda; $\chi^2_{20} = 49.4;\ P = 0.0003$) and accounted for 91 % of the variation between groups (Figure 1b). The French population expressed a substantially less pronounced behavioural gregarisation relative to the Malagasy population (test on Mahalanobis distances: $P = 0.0122$ and $P < 0.0001$ respectively). Moreover, contrary to the French population, behavioural changes induced by crowding history in the Malagasy population were typified by an increase in activity, consistent with previous analyses of locust behaviour [27]. Malagasy locusts indeed exhibited density history-dependent increases in turn frequency and decreases in climb time and staring frequency, both being static behaviours, whereas French gregarious locusts increase their staring frequency and decrease their track speed (forward stepwise selection procedures with $F = 2.706$ on groups with histories of isolation and crowding from Madagascar and France, respectively).

Altogether, these results pointed to the conclusion that the historically outbreaking population from Madagascar changes more pronouncedly its phase when submitted to a crowding-history than the historically non-outbreaking population from France.

## Discussion

The most comprehensive theory of locust phase change at the moment is that the switch towards gregarisation primarily depends on environmental conditions, in particular population density [20]. In this study, we found evidence that a (historically) outbreaking population of *L. migratoria* expressed larger behavioural and morphometric phase changes than a related (historically) non-outbreaking population. Because both populations were raised in a common environment during our experiment, an adaptive genetically-based process is likely to be responsible for this variation in the propensity to gregarise. We found "baseline' morphometrical and behavioural differences between French and Malagasy *L. migratoria* populations. Because of the inter-population differences in the solitarious phenotypes, one could invoke that the French population expresses its gregarisation differently than other populations. The phase traits we used are based on field observations and have been widely used in previous studies [13, 26]. It is therefore more parsimonious to assume that non-outbreaking populations express smaller phase change than outbreaking populations. More population samples would have strengthened our conclusions. Unfortunately, it proved logistically impossible to obtain large enough samples from more populations of each outbreaking history at the same time. It was important for our experiment to use samples isolated from the field at exactly the same time, such that they have the same rearing history prior to the treatments, precisely because we wanted to test for the effect or rearing history. Because of the lack of replication at the population level, generalising our results to the entire *L. migratoria* species should not be done without caution. However, our results show unambiguously that a genetic determinism must be responsible for the variation in the propensity to gregarise of at least some populations.

Our experimental design indicated a genetic difference between non-outbreaking and outbreaking populations in the expression of parentally inherited-gregarisation. This genetic variation for parental effects in phase change observed among populations was highly significant, despite the fact that our experimental procedure is conservative in revealing such parental effects. As a matter of fact, measured individuals of both phases were reared in identical isolation conditions, which can only decrease the differences between the treatment groups. Environmental parental effects may hence have substantially contributed to differentiation among populations in *L. migratoria* phase change. It remains to determine the physiological factors underlying the population-level variation in the expression of parental effects. The parental effects involved in the change from the solitarious to the gregarious form were shown to be transferred in a dose-dependent way by a water-soluble factor added to the egg foam in *Schistocerca gregaria* [28]. Assuming similar physiological cues in *L. migratoria*, non-outbreaking mothers may secrete smaller quantities of, or less effective pheromones than outbreaking ones, and/or non-outbreaking offspring may be less sensitive to the maternal pheromones than outbreaking ones.

We suggest here a scenario of how differences in phase change may have evolved in the two studied *L. migratoria* populations. The high propensity to gregarise of the Malagasy population may be due to selection for maintaining phase plasticity. The inter-tropical rainy front in Madagascar is responsible for large seasonal variation in locust population densities. At the beginning of the rainy season, the migratory locusts move to South-Western Madagascar following their pluviometric optimum [29]. These large movements are facilitated by the great dispersal ability of the species and the customary North-East - South-West direction of the winds [30]. These movements and the marked demographic increase after intense rains result in high

densities of locusts in favourable areas of the Malagasy South-Western coastal zone. Conversely, during the dry season, the migratory locusts disperse over the whole island [31] in search of the scarce wet patches that are left. This dilution of locusts on large areas and the numerical decrease caused by drought [29] result in populations persisting at low density. Conversely, the French population did not change its morphometry and only slightly its behaviour after two generations of crowding history. Further work is needed to determine if this observed lack of propensity to gregarise of the French population in response to two generations of parental crowding reflects a complete lack of ability to gregarise, or a shift in threshold density and/or timing triggering phase change. Assuming that gregarisation is the ancestral state for the locusts [13], the propensity to gregarise may have been partly lost in this population either because of genetic drift since selection for the gregarious character was relaxed, and/or by counter-selection if phase plasticity is costly.

This differential evolution in the ability to gregarise between populations of outbreaking and non-outbreaking areas may have an important role in swarming patterns within the species *L. migratoria*. Global changes, primarily land use change [e.g., in Indonesia: 32] but also warming and drought [e.g., in China: 33], are progressively creating new favourable environments for the migratory locust in historically non-outbreaking areas. The risk of increasing outbreak events in such areas would be closely related to the genetic ability to gregarise of the populations of interest, which depends on the amount of effective gene flow with outbreaking populations and the time since selection for the gregarious character was relaxed. Outbreaks have not been observed in Sumatra, Borneo, and New Guinea until recently, where serious outbreaks have occurred, correlated to anthropogenic changes favouring increases in population density [32]. Further research would be needed to

determine to which extent these recent outbreaks are due to selection for gregarisation among autochthonous locusts or to immigrant locusts from outbreaking areas.

The differential evolution of the gregarious character among closely related populations also provides new insight into the evolutionary diversification within the family Acrididae. The ability to change phase from solitarious to gregarious in response to population density increase has evolved multiple times within this family, resulting in the phylogenetically heterogeneous group of 'locusts' [13]. In particular, the swarming character has evolved three times within the genus *Schistocerca*, with the more frequently swarming species, *S. gregaria*, at the base of the phylogeny [34]. Behavioural propensities to gregarise among frequently and rarely swarming related *Schistocerca* species have been shown to evolve differentially [35]. Moreover, some other grasshopper species have also been shown to demonstrate some gregarious-like characteristics in crowded conditions, such as melanisation (e.g., *Schistocerca lineata* [36]), aggregation behaviour [37] or high metabolism (e.g., *Aiolopus thalassinus* [38]). This distribution of the gregarious character, which does not reflect the phylogeny of the group, is consistent with the hypothesis that genetic variation for phase change may have been present during the evolutionary past of many Acridids, but that the capacity for phase change has evolved several times via genetic assimilation [39, 40]. Through this process, reaction norm evolution can transform initially plastic phenotypes into genetically determined constitutive phenotypes [41]. Our demonstration that a *L. migratoria* population that rarely swarms lacks, at least partly, the genetic capacity to mount the appropriate density-dependent responses, lends strong support to an evolution of locust phase change via genetic assimilation. The process of genetic assimilation has been invoked to explain many instances of

phenotypic evolution, and in particular of warning coloration evolution in grasshopper species, a trait that can be density-mediated [39].

One important methodological issue arises from our results. To date, phase change has usually been analysed by comparing directly individuals grown in isolation to individuals grown under crowded conditions. In particular, this approach was used in the two previous studies attempting to document an intra-specific genetic variation in locust phase change. Results of these experimental studies were unclear and conflicting. In the non-outbreaking species *Schistocerca americana*, first-instar nymphs exhibited geographic variation in their expression of changes in behaviour with crowding, but fifth-instar nymphs did not [35]. On the other hand, in *L. migratoria*, Heifetz et al. [42] have found that a historically non-outbreaking population was not affected by crowding in its morphometry but unexpectedly more markedly in its activity level than a historically outbreaking population. Phase change is distinct from common phenotypic plasticity in that the environment of the parents is also expressed as phenotypic variation in offspring [21]. The common approach hence confounds the proximal density effects, resulting from an individual's own experience, with environmentally inherited (e.g., non-genetic parental) effects. This confusion may explain the at least partially contradictory results of previous studies. Our study hence highlights the importance of studying phase change inherited by isolation or crowding histories (i.e., that accrues across generations) instead of that expressed in isolation or crowding conditions (i.e., occuring within the lifetime of an individual) to cast off the confounding density effects, despite the fact that appropriate experimental protocols are time and manpower demanding (five generations of rearing vs. one usually). Proximal density effects may also confound the results of whole-genome expression analyses of locusts that have just come within reach in the perspective to

identify major physiological triggering processes of phase transition [43]. Kang et al. [44] identified more than 500 differentially expressed genes between laboratory induced solitarious isolated and gregarious crowded *L. migratoria* individuals. This result may suggest that molecular bases of phase change are elaborate. However, many of the gene expression differences may actually not be related to phase differences but to the different laboratory environments in which the two phases were reared. Hence, a more promising investigation axis to appropriately identify molecular inducers of locust phase change would be whole-genome comparisons of phases inherited under isolation or crowding histories to cast off the confounding proximal density effects, as done in the present study.

## Conclusions

This study demonstrates that intra-specific genetic variation is involved in locust phase change. This finding has important implications in interpreting changes of swarming patterns observed within locust species in relation to global changes as well as for our understanding of phase polyphenism evolution within the family Acrididae. The existence of a genetic variation in the reaction norm to parental environments further underlines the importance of environmental parental effects in the evolution of natural populations.

## Methods

### Populations

*L. migratoria* insects for this experiment were collected in two sites, Betioky (South-Western Madagascar) and Narbonne-plage (Southern France). The two populations are characterized by contrasted historical patterns of outbreak events. The intertropical

environment in Madagascar is cyclically favourable to increases in population density; five intense outbreaks and two incipient outbreaks controlled by insecticides have been recorded during the last century [45]. The population from Madagascar was thus considered as a historically frequently outbreaking population. In Southern France, a single outbreak, originating from the Western coast of the Black Sea, has been reported, in the XIV$^{th}$ century [46]. The population from France was therefore considered as a historically non-outbreaking population. At the moment of sampling, field densities were low (i.e., 200 and 280 adults per ha for the Malagasy and French populations respectively; results not shown), well below the critical density for first phase change manifestations (i.e., 2000 adults per ha [31]). All individuals collected from the Malagasy and French populations corresponded to the solitarious form. The locusts from Madagascar were sampled three years after the last outbreak event recorded (i.e., after approximately 12 generations assuming a generation time of four generations per year [31]). The two sampled populations were closely related as shown by genetic analysis at neutral microsatellite loci ($F_{ST}$ = 0.083; [47] and results not shown).

### Experimental design

Disentangling the phase change that accrues across generations from that occurring within the lifetime of an individual requires a close control of the density conditions of an individual across generations and homogeneous environmental conditions at the stage of measurements. The population density encountered in the field by the Malagasy and French populations might differ even if we sampled solitarious-like locusts in both cases. In particular, populations of South-Western Madagascar might experience high density conditions during the rainy season, and some gregarious populations were occasionally observed in this area. Since the parental density is

known to affect the phase characteristics across generations in locusts [28], a potential variation in the propensity to gregarise between two populations may be due to uncontrolled parental effects. To control for parental histories, both populations were reared in isolated conditions during two generations after sampling. In *Schistocerca gregaria*, the behavioural change acquired after short periods of crowding is rapidly lost [48]. Hence, if we assumed a similar time-course of behavioural phase change in *L. migratoria*, the insects were solitarious at the end of this reset step. We then reared individuals under isolated or crowded conditions for two subsequent generations. Finally, we measured phase-related traits of insects of the next generation (i.e., $5^{th}$ generation) reared in individual cages. Figure 2 presents an overall view of the experimental protocol and the resulting four groups we obtained: Malagasy population with a rearing history of isolation (*MI*), Malagasy population with a rearing history of crowding (*MC*), French population with a rearing history of isolation (*FI*), and French population with a rearing history of crowding (*FC*). To prevent that genetic diversity was depleted by drift, each generation started with 86 to 232 larvae from different egg pods (at least 14). As a result, genetic variation among the individuals within each four groups is substantial as measured by the expected heterozygosity ($0.53 \leq H_E \leq 0.80$) at neutral microsatellite loci ([47] and results not shown).

Because the measured individuals of different treatments were raised in a common environment, differences between them cannot be ascribed to differences in their environmental conditions, but only to differences in their origin (Madagascar or France) and/or their rearing history. However, phase-related traits of the measured individuals under different treatments also depend on the density conditions these individuals experienced themselves. Adults measured for morphometrics were reared

alone throughout their entire lives. Hatchlings measured for behaviour were separated from siblings on the day of emergence, and so remained together for few hours after which they were reared in isolation for the rest of their lives. In *Schistocerca gregaria*, nymphs treated by crowding for few hours and then re-isolation for only one hour were statistically similar in behaviour than untreated nymphs [48]. Assuming a similar time-course of behavioural phase change in nymphs of *L. migratoria*, contact with siblings may have only slightly affected behaviours of measured freshly hatched nymphs (1 day post hatching), and so their prevalent environment remained "isolation". Consequently, for both Malagasy and French crowding-history groups, parental experience of crowding conditions might be partly counter-balanced by the individual experience of isolation conditions. This counter-balancing effect on measured “gregarisation” makes our experimental design conservative. The fact that the Malagasy population showed a larger propensity to gregarise than the French population in spite of the potential confounding effect of rearing measured individuals in isolation strengthens our conclusions.

### Rearing

The locusts were maintained under either isolated or crowded conditions. Isolated rearing facilities were similar to those described in Hoste et al. [49], with slight modifications. Although the most potent stimulus causing solitarious locusts to assume gregarious traits is physical contact [50], the combination of visual and olfactory stimuli is also gregarising [51]. Consequently, we rendered cages opaque to visually isolate insects from each other. We also homogenized light conditions within each drawer adding ‘cool’ tube lamps every two lines and we ventilated cages through a holed top. Crowd-reared locusts were kept in a separate room in 19x19x24 cm cages at a density of about 40 individuals per cage. Different cages (at least three) were used

for each population at each generation of crowded rearing (i.e., 3rd and 4th generation; Figure 1). Those cages were ventilated and lit in a similar way to that of the individual cages. Isolated and crowded locusts were reared under completely identical room and feeding regimes. Environment rooms were under a 14 h light/10 h dark cycle, a fluctuating 14-10 h temperature regime of 32-28°C, a constant humidity of 50%, and one complete air renewal every 3 min to minimize olfactory contact among cages. Insects were fed every two days with wheat sprouts, supplemented by wheat bran for adults. Upon reaching sexual maturity, isolated females and males were placed together for 48 h to ensure insemination. Mating was done between isolated adults originating from different parents by using their reference coordinates. For both isolated and crowded conditions, egg pods were obtained in 100x50 mm diameter plastic tubes filled with moist sand (10 parts sand; 1 part water). Following oviposition, tubes were monitored for hatching which typically occurred in 12 days.

### Assaying phase state

Because phase change is a composite character, efforts to define phase state based on a single trait that changes during phase transition are illusory [52]. We considered two widely used sets of phase characteristics: morphometry, which is traditionally used in the field, and behaviour, which is widely used since the recent establishment of an assay and descriptive framework for locust behaviour [reviewed in 27].

Four morphometrical variables were considered (see Additional file 2 for illustrations of the measurements): (i) the ratio of the length of the fore wing on the length of the hind femur; (ii) the ratio of the length of the hind femur on the maximum width of the head; these two ratios are commonly used for characterizing morphometrical phase state [53]; (iii) the ratio of the maximum height of the pronotum on the length of the pronotum, which translates the well-known opposition

between the quite convex pronotum of solitarious locusts and the flat pronotum of the gregarious type; (iv) the minimum distance between the eyes, which seems to be one of the most phase discriminating phase variables [54]. We measured only adult females (40 days post-hatch) to cast off the sex-specificity of morphometrical phase change [26]. For each group (*MI*, *MC*, *FI*, *FC*), 22 to 27 insects were measured with electronic sliding callipers (accuracy 0.01 mm).

We also employed an individual-based behavioural assay developed by Roessingh et al. [55] with slight modifications. We observed 77 to 108 first-instar nymphs (1 day post-hatching) per group. Each locust was introduced via a modified syringe into the middle of a rectangular arena (30.5 cm long x 15 cm wide x 10 cm high [21]). Behind a perforated transparent plastic partition at one end of the arena and within a 4 x 15 cm backlit chamber, we placed, as a stimulus group, 80 first-instar nymphs belonging to the same population as the individual tested. Because individuals with a crowding history were not available at the time of the experiment, we used individuals with an isolation history (i.e., *MI* or *FI*), making our behavioural assay conservative. At the other end of the arena, there was a similar, but empty backlit chamber. The behaviour of the test insect was recorded on an event recorder in real time for 3 mn after introduction into the arena. We calculated, from the raw behavioural records for each test locust, values for eleven variables describing the attraction/repulsion of the insect to the backlit chambers of the arena and to the stimulus chamber separately, the tortuosity and speed of the locust track during the assay, some locomotory events, and the incidence of swaying (see Additional file 2 for a list of all eleven variables).

### Statistical analyses

Because measured morphometric and behavioural traits have been identified as key traits in *L. migratoria* phase change and quantified [26], we used them simultaneously in two multivariate analyses of morphometric and behavioural phases respectively. To determine the effects of the interaction Populational origin x Rearing history on *L. migratoria* morphometry and behaviour, we employed two full factorial MANOVAs on morphometric and behavioural data separately, with Populational origin and Rearing history as fixed factors. When appropriate, variables were Box-Cox transformed to conform to MANOVA assumptions. We then used a canonical discriminant analysis to compare groups belonging to the same population of origin, i.e. Madagascar or France, but with histories of isolation and crowding through pairwise tests on Mahalanobis distances and plots of the canonical functions. Statistical analyses were performed with the STATISTICA package v.6.1. (http://www.statsoft.com).

## Authors' contributions

The initial idea of the study was conceived by AE, ML, and YM. AE, MPC, and YM designed the experimental protocol. AF and MPC constructed the rearing racks and cages. ML provided the Malagasy insects. AF and MPC provided the French insects. AAS, AF, and MPC reared the locusts. AAS and MPC collected the morphometrical and behavioural measures. MPC carried out the statistical analyses and drafted the manuscript with substantial help from YM. AE gave valuable comments on the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We would like to thank B. Hoste for having trained M.P.C. in behavioural measures at the Catholic University of Leuven (Belgium) and P. Roessingh (University of Amsterdam, Nederland) for having provided the source code of the behavioural event recorder program. We also thank S. Piry (Centre de Biologie et de Gestion des Populations, France) who provided some modifications to the source code of the later program. We also benefited from constructive comments on an earlier version of the manuscript from N. Charbonnel (Centre de Biologie et de Gestion des Populations, France) and S.J. Simpson (University of Sydney, Australia). M.P.C. was supported by a grant from the Centre National de la Recherche Scientifique. This work was partly funded by the Institut National de la Recherche Agronomique (Santé des Plantes et Environnement) and by the Centre de coopération internationale en recherche agronomique pour le développement. YM acknowledges support from the Centre National de la Recherche Scientifique and the Institut de la Recherche et du Développement.

## References

1. Schlichting CD, Pigliucci M: ***Phenotypic evolution: a reaction norm perspective***. Edited by Sinauer Associates. Sunderland, 1998.
2. West-Eberhard MJ: *Developmental plasticity and evolution*. Edited by Oxford University Press. Oxford, 2003.
3. Laurila A, Karttunen S, Merilä J: **Adaptive phenotypic plasticity and genetics of larval life histories in two *Rana temporaria* populations**. *Evolution* 2002, **56**:617-627.

4. Michimae H: **Differentiated phenotypic plasticity in larvae of the cannibalistic salamander *Hynobius retardatus***. *Behav Ecol Sociobiol* 2006, **60**:205–211.
5. Scheiner SM: **Genetics and evolution of phenotypic plasticity**. *Ann Rev Ecol Syst* 1993, **24**:35-68.
6. Via S, Lande R: **Genotype-environment interaction and the evolution of phenotypic plasticity**. *Evolution* 1985, **39**:505-522.
7. Schmitt J, Niles J, Wulff RD: **Norms of reaction of seed traits to maternal environments in *Plantago lanceolata***. *Am Nat* 1992, **139**:451-466.
8. Mousseau TA, Dingle H: **Maternal effects in insects: examples, constraints, and geographic variation**. In: *The Unity of Evolutionary Biology: Proceedings of the ICSEB IV Meeting*. Edited by Dudley DC. Portland, 1991, 745–761.
9. Fox CW, Mousseau TA: **Maternal effects as adaptations for transgenerational phenotypic plasticity in insects**. In: *Maternal effects as adaptations.* Edited by Mousseau TA, Fox CW. Oxford, 1998, 159-177.
10. Hereford J, Moriuchi KS: **Variation among populations of *Diodia teres* (Rubiaceae) in environmental maternal effects**. 2005, **18**:124-131.
11. Applebaum SW, Heifetz Y: **Density-dependent physiological phase in insects**. *Ann Rev Entom* 1999, **44**:317-341.
12. Uvarov BP: **A revision of the genus *Locusta* L, with a new theory as to the periodicity and migrations of locusts**. *Bull Entomol Res*1921, **12**:135-163.
13. Uvarov BP: *Grasshoppers and Locusts Vol. 1*. Edited by Cambridge University Press. London, 1966.

14. Pener MP: **Locust phase polymorphism and its endocrine relations**. *Adv Insect Physiol* 1991, **23**:1-79.
15. Ellis PE: **The gregarious behaviour of marching *Locusta migratoria* migratorioides (R. & F.) hoppers**. *J Exp Biol* 1953, **30(2)**:214-234.
16. Simpson SJ, Sword GA, De Loof A: **Advances, controversies and consensus in locust phase polyphenism research**. *J Orthopt Res* 2005, **14**:213-222.
17. Wilson K, Thomas M, Blanford S, Doggett M, Simpson SJ, Moore SL: **Coping with crowds: Density-dependent disease resistance in desert locusts**. *Proc Natl Acad Sci USA* 2002, **99**:5471-5475.
18. Simpson SJ, Raubenheimer D, Behmer ST, Whitworth A, Wright GA: **A comparison of nutritional regulation in solitarious- and gregarious-phase nymphs of the desert locust, *Schistocerca gregaria***. *J Exp Biol* 2002, **205**:121-129.
19. Seidelmann K, Warnstorff K, Ferenz H-J: **Phenylacetonitrile is a male specific repellent in gregarious desert locusts, *Schistocerca gregaria***. *Chemoecol* 2005, **15**:37-43.
20. Breuer M, Hoste B, De Loof A: **The endocrine control of phase transition: some new aspects**. *Physiol Entomol* 2003, **28**:3-10.
21. Islam MS, Roessingh P, Simpson SJ, McCaffery AR: **Parental effects on the behaviour and colouration of nymphs of the desert locust *Schistocerca gregaria***. *J Insect Physiol* 1994, **40**:173-181.
22. Lauga J, Hatté M: **L'activité grégarisante du sable de ponte chez *Locusta migratoria* L. : action sur le comportement et la reproduction des individus**. *Ann Sci Nat* 1978, **20**:37-52.

23. Centre for Overseas Pest Research: *The locust and grasshopper agricultural manual*. London, 1982.

24. Uvarov BP, Zolotarevsky BN: **Phases of locusts and their interrelations**. *Bull Entomol Res*1929, **27**(1):87-90.

25. Uvarov BP, Hamilton AG: **Phase variation and rate of development in the algerian race of the migratory locust (*Locusta migratoria*, L.)**. *Bull Entomol Res*1936, **27(1)**:87-90.

26. Hoste B, Simpson SJ, Tanaka S, De Loof A, Breuer M: **A comparison of phase-related shifts in behavior and morphometrics of an albino strain, deficient in [His7]-corazonin, and a normally colored *Locusta migratoria* strain**. *J Insect Physiol* 2002, **48**:791-801.

27. Simpson SJ, McCaffery AR, Hägele BF: **A behavioural analysis of phase change in the desert locust**. *Biol Rev* 1999, **74**:461-480.

28. McCaffery AR, Simpson SJ, Islam MS, Roessingh P: **A gregarizing factor present in the egg poad foam of the desert locust *Schistocerca gregaria***. *J Exp Biol* 1998, **201**:347-363.

29. Launois M: **Influence du facteur pluviométrique sur l'évolution saisonnière du criquet migrateur, *Locusta migratoria capito* (Saussure), en phase solitaire et sur sa grégarisation à Madagascar**. *Thèse de Doctorat d'Etat ès Science Ministère de la Coopération.* Paris, 1974.

30. Duranton JF, Launois M, Launois-Luong MH, Lecoq M: **Les voies privilégiées de déplacement du criquet migrateur malgache en phase solitaire**. *Bull d'Ecol* 1979, **10**(2):107-123.

31. Lecoq M: **Les déplacements par vol du criquet migrateur malgache en phase solitaire : leur importance sur la dynamique des populations et la**

**grégarisation**. *Thèse de Doctorat d'Etat ès Science Ministère de la Coopération.* Paris, 1975.

32. Lecoq M, Sukirno: **Drought and exceptional outbreak of the oriental migratory locust, *Locusta migratoria manilensis* (Meyen 1835) in Indonesia (Orthoptera: Acrididae)**. *J Orthopt Res* 1999, **8**:153-161.
33. Zhang Z, Li D: **A possible relationship between outbreaks of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China and the El Nino episodes**. *Ecol Res* 1999, **14**:267-270.
34. Lovejoy NR, Mullen SP, Sword GA, Chapman RF, Harrison RG: **Ancient trans-Atlantic flight explains locust biogeography: molecular phylogenetics of *Schistocerca***. *Proc R Soc Lond B* 2005, **273**:767-774.
35. Sword GA: **To be or not to be a locust? A comparative analysis of behavioral phase change in nymphs of *Schistocerca americana* and *S. gregaria***. *J Insect Physiol* 2003, **49**:709-717.
36. Sword GA: **Density-dependent warning coloration**. *Nature* 1999, **397**:217.
37. Chapman RF: **A biology of locusts**. In: *The Institute of Biology's Studies in Biology 71*. Edited by Arnold E. London, 1976.
38. Heifetz Y, Applebaum SW: **Density-dependent physiological phase in a non-migratory grasshopper, *Aiolopus thalassinus thalassinus***. *Entomol Exp et Appl* 1995, **77**:251-262.
39. Sword GA: **A role for phenotypic plasticity in the evolution of aposematism**. *Proc R Soc Lond B* 2002, **269**:1639–1644.
40. Song H: **On the origin of the desert locust *Schistocerca gregaria* (Forskål) (Orthoptera: Acrididae: Cyrtacanthacridinae)**. *Proc R Soc Lond B* 2004, **271**:1641–1648.

41. Flatt T: **The evolutionary genetics of canalization**. *Quart Rev Biol* 2005, **80**:287-316.

42. Heifetz Y, Applebaum SW, Popov GB: **Phase characteristics of the Israëli population of the migratory locust, *Locusta migratoria* (L.) (Orthoptera: Acrididae)**. *J Orthopt Res* 1994, **2**:15-20.

43. De Loof A, Claeys I, Simonet G, Verleyen P, Vandersmissen T, Sas P, Huybrechts J: **Molecular markers of phase transition in locusts**. *Insect Sci* 2006, **13**:3-12.

44. Kang L, Chen X, Zhou Y, Liu B, Zheng W, Li R, Wang J, Yu J: **The analysis of large-scale gene expresion correlated to the phase changes of the migratory locust**. *Proc Natl Acad Sci USA* 2004, **101**(51):17611-17615.

45. Randriatmanantsoa M: *Manuel sur la lutte antiacridienne*. Edited by Direction de la protection des végétaux et Deutsche Gesellschaft für Technische Zusammenarbeit GmbH. Antananarivo, 1998.

46. Waloff ZV: **The distribution and migrations of *Locusta* in europe**. *Bull Entomol Res*1940, **31**(3):211-246.

47. Chapuis M-P, Loiseau A, Michalakis Y, Lecoq M, Estoup A: **Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust *Locusta migratoria***. *Mol Ecol Notes* 2005, **5**:554-557.

48. Roessingh P, Simpson SJ: **The time-course of behavioural change in nymphs of the desert locust, *Schistocerca gregaria***. *Physiol Entomol* 1994, **19**:191-197.

49. Hoste B, Luyten L, Claeys I, Clynen E, Rahman MM, De Loof A, Breuer M: **An improved breeding method for solitarious locusts**. *Entomol Exp et Appl* 2003, **104**(8):281-288.

50. Simpson SJ, Despland E, Hägele BF, Dodgson T: **Gregarious behavior in desert locusts is evoked by touching their back legs**. *Proc Natl Acad Sci USA* 2001, **98**(7):3895-3897.
51. Roessingh P, Bouaïchi A, Simpson SJ: **Effects of sensory stimuli on the behavioural phase state of the desert locust, *Schistocerca gregaria***. *J Insect Physiol* 1998, **44**:883-893.
52. Pener MP, Yerushalmi Y: **The physiology of locust phase polymorphism: an update**. *J Insect Physiol* 1998, **44**:365-377.
53. Dirsh VM: **Morphometrical studies on phases of the desert locust**. *Anti Locust Bull* 1953, **16**:1-34.
54. Lauga J: **Le problème de la mesure de la phase chez les acridiens migrateurs : historique et définition d'échelles phasaires chez *Locusta migratoria* L. (Insecte, orthoptère)**. *Arch Zool Exp Gén* 1977, **118**:247-272.
55. Roessingh P, Simpson SJ, James S: **Analysis of phase-related changes in behaviour of desert locust nymphs**. *Proc R Soc Lond B* 1993, **252**:43-49.

## Figures

**Figure 1 - First two canonical functions for the four groups of locusts obtained by canonical discriminant analysis of the measurements of four morphometric variables (a) and eleven behavioural variables (b).**

Note: *MI*, Locusts from Madagascar with an experimental rearing history of isolation; *MC*, Locusts from Madagascar with an experimental rearing history of crowding; *FI*, Locusts from France with an experimental rearing history of isolation; *FC*, Locusts from France with an experimental rearing history of crowding; *CF1*, First canonical

function; *CF2*, Second canonical function. The ellipses correspond to 95% confidence ellipses. Arrows show the significant distances between the centroids of groups of the same populational origin with experimental rearing histories of isolation and crowding as well as directions of gregarisation.

**Figure 2 - Experimental design and locust groups obtained.**

*M*, Individuals originating from Madagascar; *F*, Individuals originating from France; *MI*, Individuals originating from Madagascar with an experimental rearing history of isolation; *MC*, Individuals originating from Madagascar with an experimental rearing history of crowding; *FI*, Individuals originating from France with an experimental rearing history of isolation; *FC*, Individuals originating from France with an experimental rearing history of crowding; $G_0$, Generation of sampling; $G_1$ to $G_5$, First to fifth generation of experimental rearing. The numbers of larvae used for initiate each generation and treatment are shown in italicized characters. Below each resulting group (*MI*, *MC*, *FI*, and *FC*) are indicated the numbers of insects measured for morphometry (first line) and behaviour (second line). See text for more details.

# Additional files

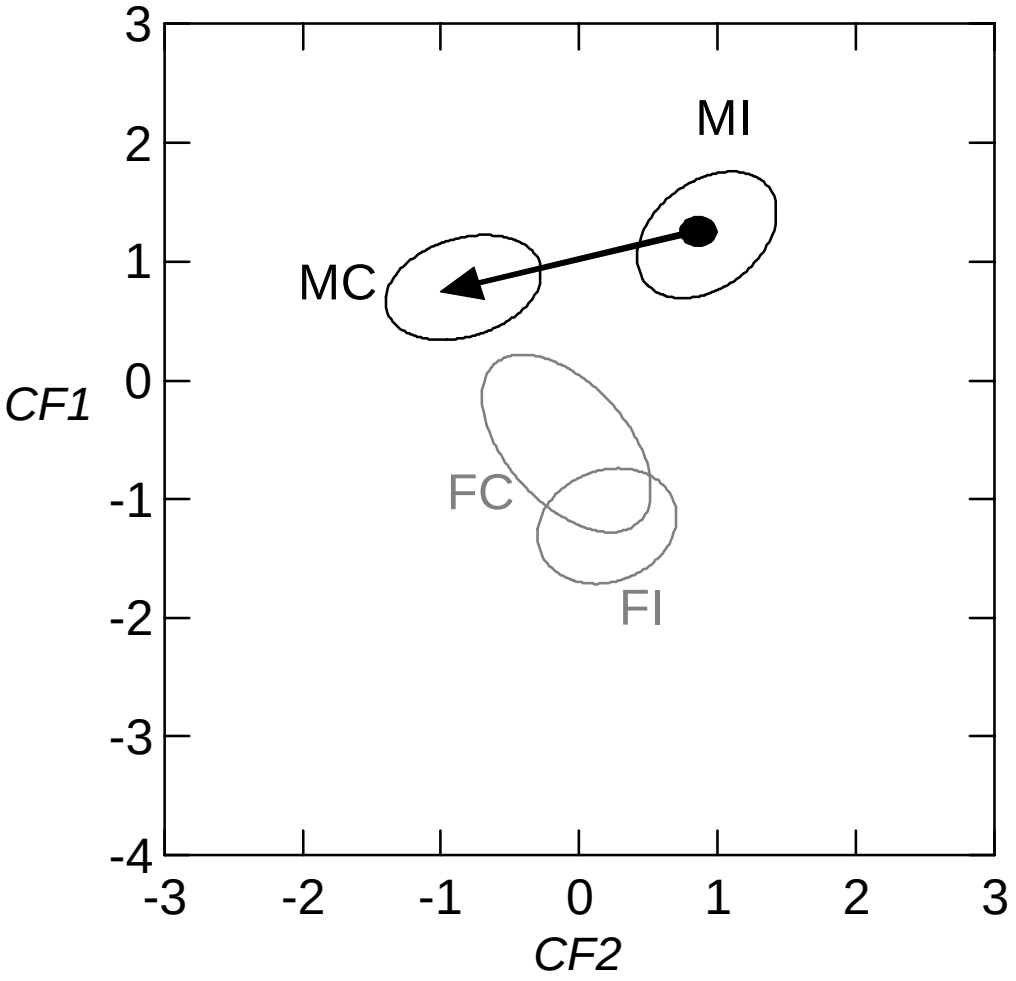
**Additional file 1 – MANOVAs of the effects of Populational origin, Rearing history and their interaction on the overall morphometry and on the overall behaviour of the four groups of locusts.**

The four MANOVA test statistics Wilks'Λ, Pillai's trace, Hotelling-Lawley trace and Roy's maximum root concluded that $H_0$ is rejected for each source of variation for both morphometric and behavioural data. The results presented in the table are those obtained with Wilks'Λ test statistic.

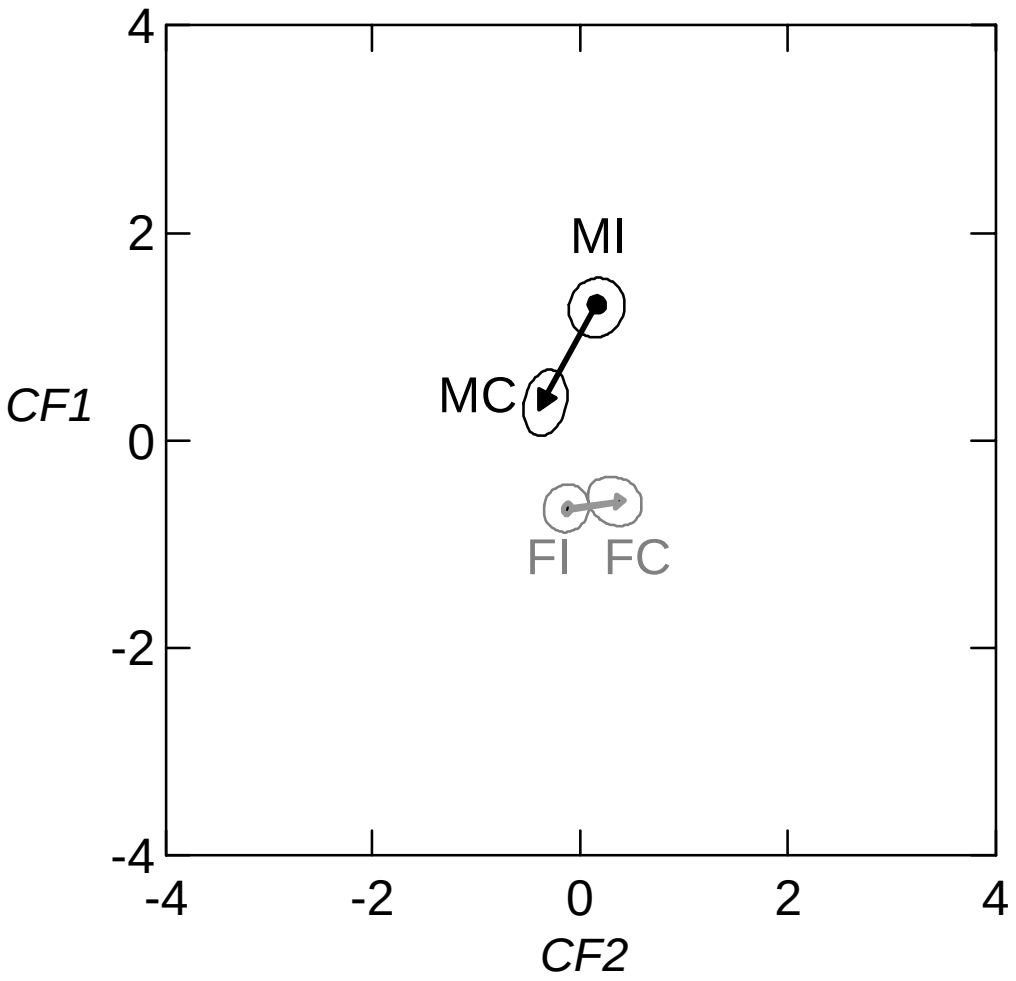
**Additional file 2 – Illustration of the measurements used for calculating the four morphometric variables (from Dirsh 1953) (a) and list of the eleven behavioural variables (b).**

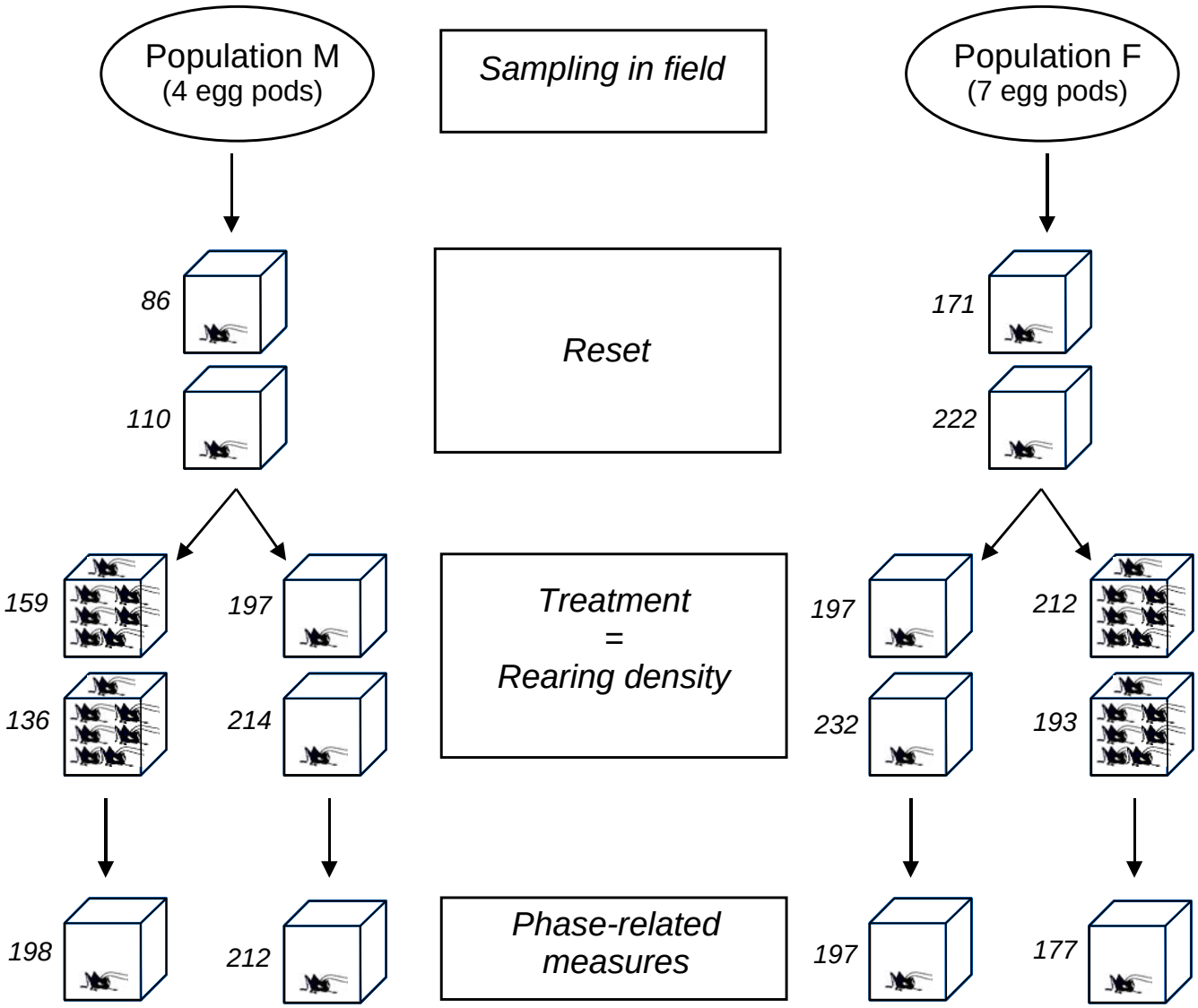
*E*, the length of the fore wing; *F*, the length of the hind femur; *C*, the maximum width of the head; *H*, the maximum height of the pronotum; *P*, the length of the pronotum; *V*, the minimum distance between the eyes.

(a) Morphometric variables



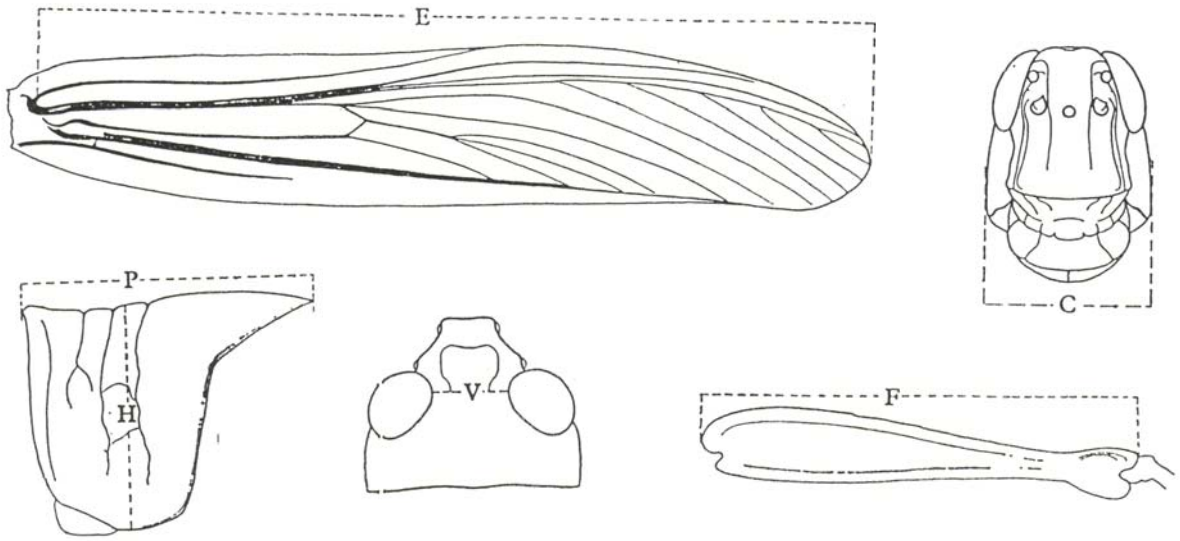

(b) Behavioural variables




Population M
(4 egg pods)
Sampling in field
Population F
(7 egg pods)
86
110
Reset
171
222
159
136
197
214
Treatment
=
Rearing density
197
232
212
193
198
212
Phase-related
measures
197
177
MC
MI
FI
FC
26
91
23
77
27
108
22
101
Rearing conditions of isolation
(1 individual per 8x8x12 cm cage)
Rearing conditions of crowding
(about 40 individuals per 19x19x24 cm cage)

| Data | Source | *DF* | Wilks'Λ | *F* | *P* |
|---|---|---|---|---|---|
| Morphometry | Populational origin | 4 | 0.5182 | 20.5 | < 0.0001 |
| | Rearing history | 4 | 0.7750 | 6.4 | 0.0001 |
| | Populational origin x Rearing history | 4 | 0.8090 | 5.2 | 0.0008 |
| | Residual | 88 | | | |
| Behaviour | Populational origin | 11 | 0.6435 | 16.8 | < 0.0001 |
| | Rearing history | 11 | 0.9060 | 3.1 | 0.0005 |
| | Populational origin x Rearing history | 11 | 0.8740 | 4.2 | < 0.0001 |
| | Residual | 331 | | | |

**(a)**




**(b)**

| | |
|---|---|
| $X_d$ | the final x-coordinate position of the insect in the arena |
| $\%_S$ | the percentage of the time spent on the stimulus side of the arena (1/3 of total surface) |
| $\%_W$ | the percentage of the time spent near the lighted walls of the arena (1/7 of total surface) |
| *St* | the track straightness |
| *Sp* | the track speed |
| *A* | the mean track angle |
| *T/t* | the turns per time |
| *W* | the walking time |
| *J* | the jumping frequency |
| *C* | the climbing time |
| *S* | the swaying frequency |

## MANUSCRIT 6 – PARENTAL CROWDING INFLUENCES LIFE-HISTORY TRAITS IN *LOCUSTA MIGRATORIA*

Marie-Pierre Chapuis, Laurent Crespin, Arnaud Estoup, Arnaud Augé-Sabatier, Antoine Foucart, Michel Lecoq et Yannis Michalakis

Prêt à soumettre à Ecological Entomology.

# Parental crowding influences life-history traits in *Locusta migratoria*

MARIE-PIERRE CHAPUIS[1,2,3], LAURENT CRESPIN[1], ARNAUD ESTOUP[1], ARNAUD AUGE-SABATIER[1], ANTOINE FOUCART[3], MICHEL LECOQ[3] and YANNIS MICHALAKIS[2]

[1]Centre de Biologie et de Gestion des Populations, Unité Mixte de Recherche Institut National de la Recherche Agronomique/Institut de Recherche pour le Développement, Campus International de Baillarguet CS 30016, 34988 Montferrier / Lez, France

[2]Génétique et Evolution des Maladies Infectieuses, UMR 2724 CNRS-IRD, IRD, 911 avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France

[3]Centre de coopération internationale en recherche agronomique pour le développement, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

**Correspondence:**

Marie-Pierre CHAPUIS
Centre de Biologie et de Gestion des Populations
Institut National de la Recherche Agronomique
Campus International de Baillarguet CS 30 016
34988 Montferrier / Lez, FRANCE
Phone: +33 (0)4.99.62.33.31
Fax: +33 (0)4.99.62.33.45
E-mail: chapuimp@ensam.inra.fr

**Running head:** Phase-specific life-histories in a locust

**Abstract.** 1. Parental environments could play an important role in controlling insect outbreaks, provided they influence changes in physiological, developmental, or behavioural life-history traits related to fluctuations in population density.

2. However, the potential implication of parental influence in density-related changes in life-history traits remains unclear in many insects that exhibit fluctuating population dynamics, particularly locusts.

3. In this study, we report a laboratory experiment which enabled us to characterize the life-history trait modifications induced by parental crowding individuals from a frequently outbreaking population of *Locusta migratoria*.

4. We found that females reproduced earlier and produced larger offspring with parental crowding conditions.

5. We discuss the evolutionary implications of these density-dependent parental effects on the life-history of *L. migratoria*.

**Keywords**. Controlled laboratory experiment, crowding, density-dependent polyphenism, life-history, locust, offspring quality, parental effects, outbreak, pest, reproductive timing.

## Introduction

The migratory locust, *Locusta migratoria*, is an agricultural pest of major importance in large areas of the Ancient world (COPR, 1982). Most of the time, the species exists at low densities in the solitarious phase, characterized by relatively cryptic and scattered individuals. During outbreaking periods, it displays huge increases of population densities with actively aggregating and swarming individuals typical of the high-density gregarious phase (Uvarov, 1966). It has been proposed that gregarious populations achieve higher growth rates than solitarious populations because they encounter more favourable environments. For instance, in Africa and Madagascar, marked numerical population increases at the beginning of outbreaks have been associated with above-average rainfall (Davey, 1956; Waloff, 1962; Dempster, 1963; Têtefort & Wintrebert, 1966; Roffey & Popov, 1968). By stimulating vegetation growth, such conditions result in abundant and high quality food resources for breeding adults. Moreover, higher soil humidity allows successive waves of females to lay and decreases egg desiccation and first-instar larval mortality. The absence of egg and larval parasites and predators has also been proposed as a factor favouring the production of locust outbreaks (Roffey & Popov, 1968; Wintrebert, 1970; Farrow, 1974). However, the potential for gregarious populations to maintain a high growth rate following outbreak initiation is often counteracted by their frequent extension into new environments that are unfavourable for sustained reproduction (Batten, 1967).

The higher growth rate of gregarious populations may also result from intrinsic differences in survival and/or fecundity between the gregarious phase individuals produced by high density conditions and the solitarious phase individuals generated under low density conditions (Uvarov, 1966). As a causal factor of fluctuations in population growth rates, a density-related shift in life-history traits could arise from genetic (Chitty, 1967) or non-genetic (Rossiter, 1996; Ginzburg, 1998) parental effects. Because of the strength of selection required to produce a major change in gene frequency in just one or a few generations, the

hypothesis of a genetic causation of short-term fluctuations in population growth rates has received so far little credit (Mitter & Schneider, 1987; but see Baltensweiler, 1984). On the other hand, a considerable amount of theoretical work indicates that the delayed effects of parental environments are capable of profoundly altering the life-history traits of insect populations (e.g., Rossiter, 1992), and are likely to generate population outbreaks (e.g., Ginzburg & Taneyhill, 1994; Rossiter, 1994). Although parental density conditions have been shown to drive behavioural and morphological phase changes in locusts (Islam *et al.*, 1994) and in *L. migratoria* in particular (Chapuis *et al.*, submitted), empirical evidence for parental effects on the density-related changes in life-history traits of their offspring is still lacking in these species (but see Albrecht *et al.*, 1959). To adequately test the hypothesis of parentally-mediated variation in life-history traits among gregarious and solitarious locusts, multigenerational studies controlling for population density, life-history traits, and environmental quality are compulsory (Rossiter, 1994; Hunter, 2002). Because *L. migratoria* is a highly mobile insect, especially in tropical regions where it displays seasonal migrations, such studies are very difficult to conduct in the field. An alternative approach consists of testing for differences in the life-history traits among the density-related phases by comparing laboratory-induced solitarious and gregarious individuals of the same locust population sample under controlled density conditions.

In the present study, we tested the prediction that gregarious females expressed life-history traits that contributed to higher population growth rates than solitarious females under homogeneous density conditions in *L. migratoria.* To this aim, we assessed the influence of solitarious and gregarious parentally-inherited phases on development, survival, and fecundity of females, as well as the quality and sex-ratio of their offspring. We compared two treatment groups of locusts with a known history of controlled isolated and crowded rearing conditions. Life-history traits may not only vary because of genetic and environmental factors acting on previous generations, but also because of the proximal environmental conditions experienced

by the individuals under study. Several previous attempts to characterise differences between gregarious and solitarious locusts failed to disentangle the genetic/environmental effects acting on previous generations from proximal environmental effects. Our experimental design, by rearing the focal experimental insects from the isolated and crowded treatment groups in the same environment (isolation), manages to cast off the confounding proximal density effects. Thus, any differences among treatments can only be ascribed to the rearing history. We used a population sample from Madagascar, an area characterized by frequent outbreaking events, and shown in a previous study to express density-dependent behavioural and morphometric phase changes (Chapuis *et al.*, submitted). We demonstrate that females respond to parental crowding conditions by reproducing earlier and producing larger offspring. Such delayed effects of parental crowding, whereby gregarious females invest more in reproduction, may have important implications for the control of *L. migratoria* outbreaks.

## Materials and Methods

*Insects and Experimental design*

Pre-reproductive adults of *L. migratoria* were collected in Betioky in South-Western Madagascar. During the $20^{th}$ century, five intense plagues and two incipient outbreaks controlled by insecticides of highly gregarious locusts have been observed in Madagascar, including the Betioky area (Randriamanantsoa, 1998). When sampling, field densities were low (ca. 200 individuals per ha; data not shown), well below the critical density for first phase change manifestations (ca. 2000 individuals per ha; Lecoq, 1975). All individuals collected corresponded to the solitarious morphological phase and were sampled three years after the last outbreak event recorded (i.e., after 12 generations assuming a generation time of four generations per year; Lecoq, 1975). Populations of South-Western Madagascar might however experience high density conditions during the rainy season, and some gregarious populations were occasionally observed in this area, even during remission periods. To

control for parental histories, individuals were reared in the laboratory under isolated conditions for two generations after sampling. In *Schistocerca gregaria*, the behavioural change acquired after short periods of crowding is rapidly lost (Roessingh & Simpson, 1994). Hence, if we assumed a similar time-course of behavioural phase change in *L. migratoria*, the insects were solitarious at the end of this reset step. We then reared individuals under isolated or crowded conditions for two subsequent generations. To prevent differentiation by genetic drift between isolated and crowded treatment groups, each generation started with 136 to 232 larvae from different egg pods (at least 14). Finally, we measured development, survival, fecundity, offspring quality, and offspring sex-ratio of female individuals of the next generation (i.e., $5^{th}$ generation) reared in individual cages. By doing so, we were able to distinguish the effects of rearing history from those of the environment of the assay. Figure 1 presents an overall view of the experimental protocol and the resulting two groups we obtained, namely individuals with a rearing history of isolation (hereafter referred to as *I*) and individuals with a rearing history of crowding (*C*). Locusts with isolation and crowding histories developed highly contrasted behavioural and morphometrical phenotypes, indicating that the protocol used here efficiently induced density-dependent phase changes in this population (Chapuis *et al.*, submitted). Rearing under isolation and crowded conditions was performed as in Chapuis *et al.* (submitted).

*Development and Reproduction analysis*

We compared development and reproduction of females of the $5^{th}$ generation with crowding (*C*) and isolation (*I*) histories. The individual development parameters were developmental time of larvae (as measured by the adult moulting day minus the hatching day) and femur length of 40 days post-hatching adults (i.e., when the cuticle had hardened). All reproductive parameters were measured for the first laid egg pod only, as in natural solitarious populations of *L. migratoria* in South-Western Madagascar most females lay only one pod (average

number of laying events per female = 1.35; Lecoq, 1975). The single exception was the time necessary for the females to lay their second egg pod after first laying (i.e., second laying day minus first laying day). This reproductive time was monitored instead of the time necessary for the females to lay their first egg pod once mated (i.e., first laying day minus mating day) because mating day was controlled by the experimenter, who placed together isolated females and males once they reached sexual maturity (see Chapuis *et al.*, submitted). The timing of reproduction was also assessed by monitoring the time necessary for the eggs to hatch (i.e., hatching day minus laying day). Moreover, we counted the number of laid eggs and estimated hatching rate (i.e., number of hatched eggs divided by number of laid eggs). Finally, we measured femur lengths of freshly hatched offspring (<24h) following Dirsh (1953), by using scanning and image analysis tools (UTSCHA, Image Tool version 3.0, developed by the University of Texas Health Science Center, San Antonio, TX, USA). Offspring sex-ratio was assessed by examining external genitalia of freshly hatched larvae (<24h) following Dirsh (1950). Egg-pods in which more than 90% of eggs died were excluded from the analysis, assuming laboratory conditions induced massive abortion or females were not successfully fertilized.

Time data, adult femur lengths, and egg numbers were analysed using ANOVAs with *Density history* as a fixed factor, using the JMP package (SAS Institute, 1995, Carry, NC, USA). Time data were logarithmically-transformed and adult femur lengths were squared-transformed to satisfy analysis of variance requirements. For proportions (hatching rate and sex-ratio), analyses of deviance were performed assuming a logit link function and a binomial error (Crawley, 1993), and using the GLIM package (Baker & Nelder, 1985). Overdispersion was accounted for following Crawley (1993). Because of unbalanced data, we analysed offspring femur lengths using a mixed model with *Density history* and *Sex* as fixed factors and *Mother* as a random factor, using the SAS system (Littell *et al.*, 1996). Because of our experimental design (see '*Discussion*' section) and the small numbers of observations (see

Table 1), we expected a low statistical power in detecting effects between *I* and *C* groups. We hence did not apply adjustment for multiple testing, such as the sequential Bonferroni (Rice, 1989), which is often considered as a too conservative procedure (Moran, 2003). Instead, statistical results were carefully discussed in comparison with previously published results and life-history theory (Yoccoz, 1991).

*Survival analysis*

Females of the 5$^{th}$ generation with crowding (*C*) and isolation (*I*) histories were regularly checked for survival throughout the study period. Due to logistic constraints, time schedules were different between the groups *I* and *C:* survival of group *I* was checked in 21 occasions while that of group *C* was checked in 28 occasions. Thus it was not possible to analyse both groups together. Further, the time intervals between consecutive checks varied from 2 to 9 days. To allow comparison between the two groups, all survival estimates were standardised to 24h. Separate analyses of survival between groups were carried out following Lebreton *et al.* (1992). At each check, each female was given a state based on morphological or reproductive characteristics: larva (denoted LAR) before the day of adult moulting, non reproductive adult (NRA) between the day of adult moulting and the day of first laying, and first-time reproductive adult (FRA) after the first laying event. We thus used a 3-states model to analyse this dataset (see Nichols *et al.*, 1992). Eleven parameters were necessary: three probabilities of survival, three probabilities of recapture (here set to one as no individual escaped from its cage during the experiment) and five probabilities of transition. With regards to transitions, all reverse transitions i.e. from FRA to NRA, from NRA to LAR and from FRA to LAR were set to zero.

Because the experiment was run under constant controlled conditions, we did not expect time-variation in the probabilities of transition and survival. Besides, because there were few observations as FRA (see '*Results'* section) we set the FRA survival and the

transition from NRA to FRA constant. However, we expected an age-dependence for the transition probability LAR to NRA, since young larvae have a lower probability to become NRA than old larvae. Before moulting to adults, larvae go through five different stages that were not recorded throughout the experiment (Uvarov, 1966). Because these different larval stages may have different survival rates, we tested for an age effect in the survival of larvae. We thus considered age to be a categorical variable with the last age class pooling several late age classes due to small sample sizes. Overall, we analysed eight models differing in whether parameters were age dependent or not. We also ran a random effect model on age-dependent survival parameters of larvae to estimate a mean and its variance over all ages from the best model among the previous eight models (Burnham & White, 2002). This random effect model also provided a straightforward way for testing for differences in survival of larvae between the two groups by performing a *Z* test on these estimates. The program MARK was used for model fitting (White & Burnham, 1999).

## Results

*F*-tests revealed that *Density history* significantly affected only the time between first and second layings and the offspring femur length ($F_{1,17}$ = 6.83 and $P$ = 0.018; $F_{1,20}$ = 5.70 and $P = 0.028$, respectively). Females with a crowding rearing history had the best parameters, with a half-fold decrease in the time to lay the second egg-pod after first laying and offspring that emerged larger (Table 1). For offspring femur length, comparisons between the two genders revealed a barely significant difference, with freshly hatched females larger than males ($F_{1,265}$ = 3.94 and $P$ = 0.048). There was also a significant among-mother variance component for offspring femur length (Wald test; $Z$ = 2.82 and $P$ = 0.002).

Model selection on survival data ranked the eight models similarly for both groups *I* (isolation rearing history) and *C* (crowding rearing history ) (see Appendix 1 for a summary of the model selection for groups *I* and *C*). The best model had age-dependence for survival of

larvae and transition from larva to non-reproductive adult. However, survival of non-reproductive adults was constant in the best model. Given the high number of observations of this class, the absence of age-dependence could not be ascribed to lack of data (see Appendix 2 for a summary of the number of observations in each state and the observed transitions between states in the dataset). For both the groups *I* and *C*, survival of non-reproductive adults and larvae were largely superior to the survival of reproductive adults (Table 2). The probability of transition from larva to non-reproductive adult overall increased with age for both groups. Differences in survival of larvae and non-reproductive adults between the groups *I* and *C* were minute (Table 2; results of the random effects model for larvae *I*: 0.976 ± 0.009 and *C*: 0.979 ± 0.008). Larger though still not significant differences between the groups *I* and *C* were found in survival for reproductive adults ($Z$ = -0.80 and $P$ = 0.42) and in the probability of transition from non-reproductive to reproductive adults ($Z$ = 0.74 and $P$ = 0.46).

All these results showed that a rearing history of crowding led to reduced female laying times and increased sizes of offspring but did not affect the developmental time, survival, fecundity, and the sex-ratio and the number of offspring.

## Discussion

In this study, we found that a historically frequently outbreaking Malagasy population of *L. migratoria* invested more in reproduction when gregarized, through both earlier reproduction (by reducing laying times) and higher offspring quality (by increasing sizes of offspring). These observed phase-related effects on reproduction were relatively small. However, our experimental procedure might be conservative to reveal such phase-related effects. Measured individuals from both groups *I* and *C* were reared in isolation conditions, which can only decrease the differences between the groups *I* and *C* due to non-parental effects. The significance of parental effects may also depend on the environment the offspring encounter (Bernardo, 1996; Plaistow *et al.*, 2006). To thoroughly test and measure the parental effects

on life-history traits of offspring reared under crowding conditions, the statistical analysis of a large number of cages would be required because individuals grown in the same cage are not independent. Moreover, more population samples would be needed to generalise our result to the entire *L. migratoria* species. However, it was important for our experiment to use samples isolated from the field at exactly the same time, such that they have the same rearing history prior to the treatments, precisely because we wanted to test for the effect of rearing history. Unfortunately, it proved logistically impossible to obtain large enough samples from more populations with a frequently outbreaking history at the same time.

In our experimental design, measured females of solitarious and gregarious phases were raised in a common environment (isolation conditions). Therefore, trans-generational effects induced by density must be responsible for the larger investment in reproduction observed for gregarious individuals. This result parallels that of Lauga & Hatté (1978), in which *L. migratoria* isolated females hatching in sand used many times for egg laying by crowded females laid heavier and more eggs within their reproductive lifetime than isolated females hatching in sterilized sand. Moreover, this result is, at least partly, consistent with the findings of Albrecht *et al.* (1959). These authors reared a crowded stock of *L. migratoria* individuals in isolation conditions for successive generations, and reported that the weight of eggs was smaller but the fecundity potential (ovariole number) of females larger in successive generations. The observed trans-generational elevation of population quality with high-density conditions is likely to increase gregarious population growth rates; it hence might be of particular importance in the production and/or duration of outbreaks of this species.

The evolutionary factors underlying phase-level variation in life-history strategies of *L. migratoria* have still to be identified. Life-history evolutionary theory predicts that the earlier organisms are sensitive to mortality the earlier they should invest in reproduction (reviewed in Reznick, 2002). Selection favouring plasticity in offspring quality, whereby a mother that experiences a resource-limited environment allocates more resources to her

offspring, has been demonstrated in natural insect populations. For instance, females of seed beetles lay larger eggs in response to increasing adult density – an indicator that their hatchlings will encounter severe competition for food (reviewed in Fox & Mousseau, 1998). Because gregarious populations attain huge size and invade new areas, their resource availability is expected to be lower and more unpredictable than that of solitarious populations, with a subsequent negative impact on their survival probability (Batten, 1967). It is hence possible that the observed phase-level variations in life-history strategies in *L. migratoria* have evolved through phase-level variations in age-specific mortality. We did not detect significant differences in female survival with phase state. We found that producing larger offspring did not come at the cost of producing fewer offspring either, though such trade-offs are commonly observed in natural populations (Smith & Fretwell, 1974). The apparent absence of trade-offs may be explained by the fact that such trade-offs are in general less detectable under benign conditions (Zera & Harshman, 2001). Indeed, our experiment was conducted under benign isolation conditions, since animals were not in competition for food or egg laying space and were not stressed by contact or toxic wastes of conspecifics. Moreover, laboratory conditions precluded the effects of stressful natural factors such as predators or parasitism which may differentially affect survival of individuals of the two phases.

## Acknowledgments

M.P.C. was supported by a CNRS BDI grant. This work was partly funded by INRA (SPE department) and CIRAD. YM acknowledges support from CNRS and IRD. We also benefited from constructive comments on an earlier version of the manuscript from N. Charbonnel (CBGP, France) and G. Sword (University of Sydney, Australia).

## References

Albrecht, F.O., Verdier, M., & Blackith, R.E. (1959) Maternal control of ovariole number in the progeny of the migratory locust. *Nature*, **184**, 103-104.

Baker, R.J. & Nelder, J.A. (1985) *The GLIM system, Release 3.77, Manual*. Numerical Algorithms Group, Oxford.

Baltensweiler, W. (1984) The role of environment and reproduction in the population dynamics of the larch bud moth, *Zeiraphera diniana* Gn. (Lepidoptera, Tortricidae). *Advances in Invertebrate Reproduction*, **3**, 291-302.

Batten, A. (1967) Seasonal movements of swarns of *Locusta migratoria migratorioides* (R. & F.) in western Africa in 1928 to 1931. *Bulletin of Entomological Research*, **57**, 357-380.

Bernardo, J. (1996) Maternal effects in animal ecology. *American Zoologist*, **36**, 83-105.

Burnham, K.P. & White, B.N. (2002) Evaluation of some random effects methodology applicable to bird ringing data. *Journal of Applied Statistics*, **29**, 245-264.

Chapuis, M.-P., Estoup, A., Augé-Sabatier, A., Foucart, A., Lecoq, M., & Michalakis, Y. (submitted) Genetic variation for propensity to gregarise in *Locusta migratoria*. *Evolutionary Ecology*.

Chitty, D. (1967) The natural selection of self-regulatory behaviour in animal populations. *Proceedings of the Ecological Society of Australia*, **2**, 51-78.

COPR (1982) *The Locust and Grasshopper Agricultural Manual*. Centre for Overseas Pest Research, London.

Crawley, M.J. (1993) *GLIM for ecologists*. Blackwell Scientific Publications, Oxford.

Davey, J.T. (1956) The seasonal migrations and dynamics of populations of the African migratory Locust (*Locusta migratoria migratorioides* Reiche and Fairmaire) in the outbreak area (Orth.). *Bulletin de la Société Entomologique de France*, **61**, 18-24.

Dempster, J.P. (1963) The population dynamics of grasshoppers and locusts. *Biological Reviews*, **38**, 490-529.

Dirsh, V.M. (1950) A practical table for the determination of sexes of nymphs of *Locusta migratoria migratorioides* (R & F). *Proceedings of the Royal Entomological Society of London Series B*, **19**, 136-137.

Dirsh, V.M. (1953) Morphometrical studies on phases of the desert locust. *Anti Locust Bulletin*, **16**, 1-34.

Farrow, R.A. (1974) Comparative plague dynamics of tropical *Locusta* (Orthoptera, Acrididae). *Bulletin of Entomological Research*, **64**, 401-411.

Fox, C.W. & Mousseau, T.A. (1998) Maternal effects as adaptations for transgenerational phenotypic plasticity in insects. in T.A. Mousseau and C.W. Fox, *Maternal effects as adaptations*. Oxford University Press, New-York, 159-177.

Ginzburg, L.R. (1998) Inertial growth: population dynamics based on maternal effects. in T.A. Mousseau and C.W. Fox, *Maternal effects as adaptations*. Oxford University Press, New-York, 42-53.

Ginzburg, L.R. & Taneyhill, D.E. (1994) Population cycles of forest Lepidoptera: a maternal effect hypothesis. *Journal of Animal Ecology*, **63**, 79-92.

Hunter, M.D. (2002) Maternal effects and the population dynamics of insects on plants. *Agricultural and Forest Entomology*, **4**, 1-9.

Islam, M.S., Roessingh, P., Sinervo, B., & McCaffery, A.R. (1994) Effects of population density experienced by parents during mating and oviposition on the phase of hatchlings desert locusts, *Schistocerca gregaria*. *Proceedings of Royal Society of London series B*, **257**, 93-98.

Lauga, J. & Hatté, M. (1978) L'activité grégarisante du sable de ponte chez *Locusta migratoria* L. : action sur le comportement et la reproduction des individus. *Annales des Sciences naturelles*, **20**, 37-52.

Lebreton, J.-D., Burnham, K.P., Clobert, J., & Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monography*, **62**, 67-118.

Lecoq, M. (1975) Les déplacements par vol du criquet migrateur malgache en phase solitaire : leur importance sur la dynamique des populations et la grégarisation. Thèse de Doctorat d'Etat ès Science. Ministère de la Coopération, Paris.

Littell, R.C., Milliken, G.A., Stroup, W.W., & Wolfinger, R.D. (1996) *SAS System for Mixed Models*, SAS Institute, Cary, NC.

Mitter, C. & Schneider, J.C. (1987) Genetic change and insect outbreaks. in P. Barbosa and J.C. Schultz, eds. *Insect outbreaks*. Academic Press, New York, 502-532.

Moran, M.D. (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oïkos*, **100**, 403-405.

Nichols, J.D., Sauer, J.R., Pollock, K.H., & Hestbeck, J.B. (1992) Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology*, **73**, 306-311.

Plaistow, S.J., Lapsley, C.T., & Benton, T.G. (2006) Context-dependent intergenerational effects: the interaction between past and present environments and its effect on population dynamics. *The American Naturalist*, **167**, 206-215.

Pradel, R., Wintrebert, C.M.A., & Gimenez, O. (2003) A proposal for a goodness-of-fit to the Arnason-Schwarz multisite capture-recapture model. *Biometrics*, **59**, 43-53.

Randriatmanantsoa, M. (1998) *Manuel sur la lutte antiacridienne*. Direction de la protection des végétaux et Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, Antananarivo.

Reznick, D., Bryant, M.J., & Bashey, F. (2002) *r*- and *K*-selection revisited: the role of population regulation in life-history evolution. *Ecology*, **83**, 1509-1520.

Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Roessingh, P. & Simpson, S.J. (1994) The time-course of behavioural change in nymphs of the desert locust, *Schistocerca gregaria*. *Physiological Entomology*, **19**, 191-197.

Roffey, J. & Popov, G.B. (1968) Environmental and behavioural processes in a desert locust outbreak. *Nature*, **219**, 446-450.

Rossiter, M.C. (1992) The impact of resource variation on population quality in herbivorous insects: a critical component of population dynamics. in M.D. Hunter, T. Ohgushi, and P.W. Price, eds. *Resource distribution and animal-plant interactions*. Acadamic Press, San diego, CA, 13-42.

Rossiter, M.C. (1994) Maternal effects hypothesis of herbivore outbreak. *BioScience*, **44**, 752-762.

Rossiter, M.C. (1996) The incidence and consequences of inherited environmental effects. *Annual Review of Ecology and Systematics*, **27**, 451-476.

Smith, C.C. & Fretwell, S.D. (1974) The optimal balance between size and number of offspring. *American Naturalist*, **109**, 499-506.

Têtefort, J.P. & Wintrebert, D. (1966) Effets édaphiques et biotiques de l'inversion pluviométrique des saisons, important facteur de pullulation des acridiens migrateurs dans le sud-ouest malgache. *Entomophaga*, **2**, 305-310.

Uvarov, B.P. (1966) *Grasshoppers and Locusts, Vol.1*. University Press, Cambridge.

Waloff, Z. (1962) Local outbreaks of the African migratory locust in the Red Sea area. *FAO Plant Protection Bulletin*, **10**, 61-63.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120-139.

Wintrebert, D. (1970) Identité, écologie et comportement du criquet migrateur dans le sud-ouest Malgache. *Annales de la Société Entomologique de France*, **6**, 35-152.

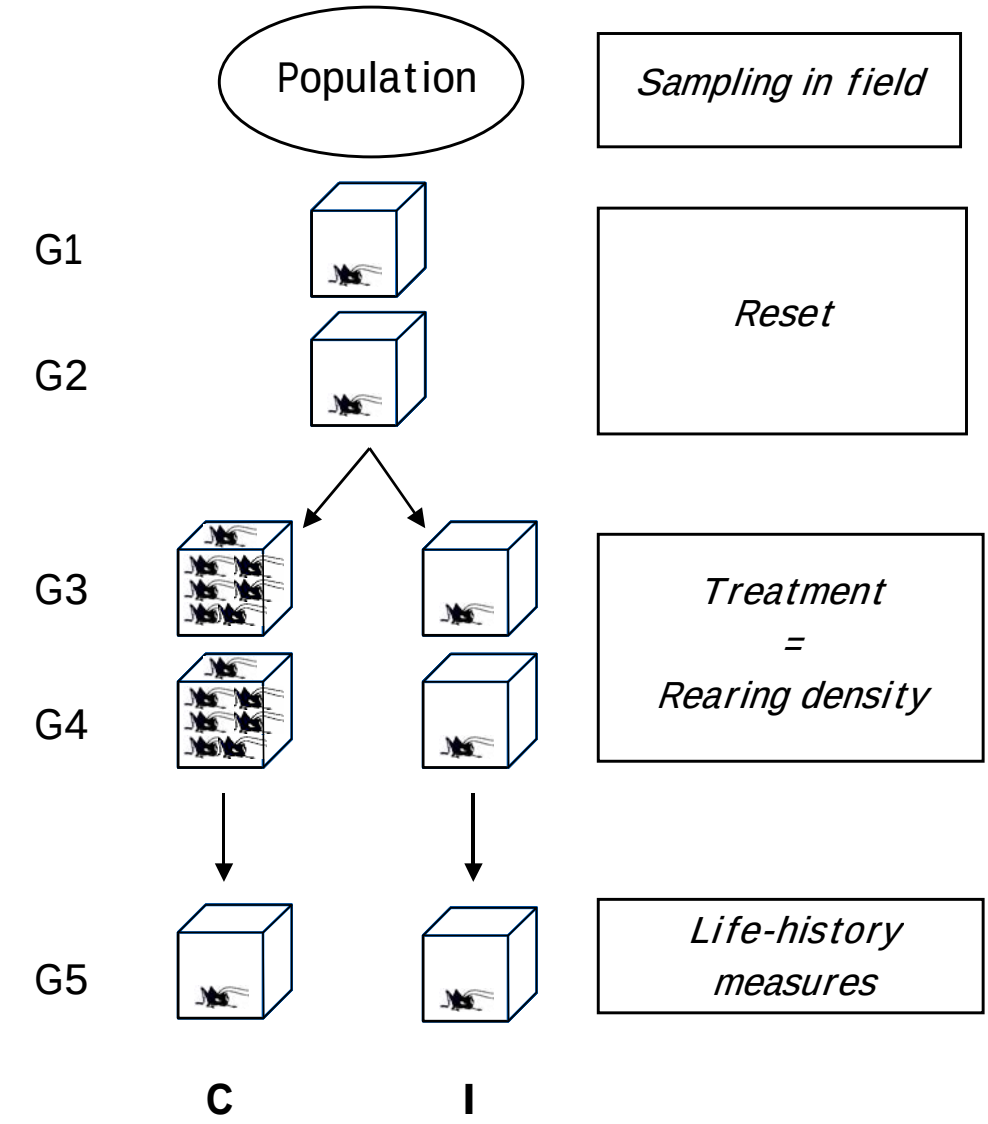
Yoccoz, N.G. (1991) Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of Ecological Society of America*, **72**, 106-111.

Zera, A.J. & Harshman, L.G. (2001) The physiology of life history trade-offs in animals. *Annual Review of Ecology and Systematics*, **32**, 95-126.

## Figure legends

**Figure 1. Experimental design.** *I*: Rearing history = isolation conditions, *C*: Rearing history = crowding conditions, *G1* to *G5*: first to fifth generation of experimental rearing. See text for more details. Details of rearing conditions are given in Chapuis *et al.* (submitted).

**Fig. 1.**


Population
Sampling in field
G1
G2
Reset
G3
G4
Treatment
=
Rearing density
G5
Life-history
measures
C
I
Rearing conditions of isolation
(1 individual per 8x8x12 cm cage)
Rearing conditions of crowding
(about 40 individuals per 19x19x24 cm cage)

## Tables

**Table 1. Summary of the *Density history* effect on development and reproductive traits in *L. migratoria*.** Means and standard errors were computed using raw (i.e., not transformed) data. Lengths are in mm and times in days. *F*-tests are given for the *Density history* effect only, but for offspring femur length data, *Sex* and *Mother* effects were also tested (as a fixed and a random factor, respectively; see text for more details). Significant differences of the means ($P < 0.05$) are shown in bold characters. *n*: Number of observations; *I*: Isolation rearing history; *C*: Crowding rearing history.

| | | *n* | | Mean ± Standard Error | | *F*-test | *P*-value |
|---|---|---|---|---|---|---|---|
| | | *I* | *C* | *I* | *C* | | |
| ***Development traits*** | | | | | | | |
| | Developmental time of larvae | 54 | 44 | 28.98 ± 0.46 | 29.63 ± 0.59 | $F_{1,96} = 0.64$ | 0.425 |
| | Adult femur length | 23 | 26 | 25.23 ± 0.25 | 24.63 ± 0.25 | $F_{1,47} = 2.87$ | 0.097 |
| ***Reproductive traits*** | | | | | | | |
| | Second laying day minus first laying day | 13 | 6 | 6.69 ± 0.50 | 4.50 ± 0.67 | $F_{1,17} = 6.83$ | **0.018** |
| First pod | Hatching day minus laying day | 12 | 8 | 12.67 ± 0.38 | 12.37 ± 0.53 | $F_{1,18} = 0.24$ | 0.627 |
| First pod | Egg number | 7 | 7 | 26.28 ± 2.38 | 28.14 ± 2.77 | $F_{1,12} = 0.26$ | 0.619 |
| First pod | Offspring femur length | 159 | 129 | 0.389 ± 0.003 | 0.400 ± 0.002 | $F_{1,20} = 5.58$ | **0.028** |
| First pod | Hatching rate | 7 | 7 | 0.649 ± 0.091 | 0.723 ± 0.080 | $F_{1,12} = 1.77$ | 0.208 |
| First pod | Sex-ratio at hatching | 7 | 7 | 0.473 ± 0.020 | 0.501 ± 0.036 | $F_{1,12} = 0.55$ | 0.472 |

**Table 2. Estimates and standard errors of survival (Φ) and transition (Ψ) probabilities under the best three-state model (i.e., model 8; see Appendix 1).** The three states are LAR: larvae; NRA: non-reproductive adult; and FRA: first-time reproductive adult. The estimates of each age class are presented for the age-dependent parameters (survival of larvae and the transition from larva to non-reproductive adult; see text for more details). Survival estimates are standardized to 24h. Program MARK does not provide estimates of standard error for estimates close to the boundaries of the parameter space (0 or 1). *I*: isolation rearing history; *C*: Crowding rearing history.

| Parameter | *I* | *C* |
|---|---|---|
| Φ(FRA) | 0.909 ± 0.023 | 0.933 ± 0.032 |
| Φ(NRA) | 0.971 ± 0.005 | 0.974 ± 0.005 |
| Φ(LAR) | 0.947 ± 0.010, 0.984 ± 0.006, 0.992 ± 0.004, 0.991 ± 0.005, 0.969 ± 0.013, 0.913 ± 0.042, 0.961 ± 0.038 | 1.000, 0.969 ± 0.010, 0.988 ± 0.007, 0.997 ± 0.003, 1.000, 0.975 ± 0.010, 0.981 ± 0.009, 0.975 ± 0.011, 0.878 ± 0.038, 0.906 ± 0.040 |
| Ψ(LAR→NRA) | 0, 0, 0.029 ± 0.021, 0.322 ± 0.059, 0.621 ± 0.080, 0.400 ± 0.154, 0.714 ± 0.170 | 0, 0, 0, 0, 0.014 ± 0.014, 0, 0.113 ± 0.040, 0.420 ± 0.070, 0.450 ± 0.111, 0.500 ± 0.144 |
| Ψ(NRA→FRA) | 0.187 ± 0.035 | 0.178 ± 0.040 |

## Appendices

**Appendix 1. Summary of the model selection for survival data of groups with (a) isolation and (b) crowding rearing histories.** For all models, the transition NRA→FRA and the survival of state FRA were set constant. The fit of the models was assessed by Akaike's Information Criterion ($AIC_c$) where a lower value shows better fit. Models are ranked by $AIC_c$ with best model shown in bold. Notations follow Lebreton et al. (1992). .: constant (intercept only model). *np* gives the number of identifiable parameters in the model under consideration. LAR: larvae; NRA: non-reproductive adult; FRA: first-time reproductive adult.

**(a)**

| Model number | Survival | | Transition | Deviance | *np* | $AIC_c$ |
|---|---|---|---|---|---|---|
| | LAR | NRA | LAR→NRA | | | |
| 1 | . | Age | . | 514.16 | 13 | 975.48 |
| 2 | . | . | . | 517.62 | 5 | 962.51 |
| 3 | Age | Age | . | 484.18 | 19 | 958.08 |
| 4 | Age | . | . | 487.64 | 11 | 944.82 |
| 5 | . | Age | Age | 387.26 | 19 | 861.17 |
| 6 | . | . | Age | 390.72 | 11 | 847.90 |
| 7* | Age | Age | Age | 357.28 | 25 | 843.98 |
| **8** | **Age** | **.** | **Age** | **360.74** | **17** | **830.43** |

**(b)**

| Model number | Survival | | Transition | Deviance | *np* | $AIC_c$ |
|---|---|---|---|---|---|---|
| | LAR | NRA | LAR→NRA | | | |
| 1 | . | Age | . | 612.26 | 15 | 906.11 |
| 2 | . | . | . | 615.86 | 5 | 889.25 |
| 3 | Age | Age | . | 556.44 | 24 | 869.12 |
| 4 | Age | . | . | 560.05 | 14 | 851.84 |
| 5 | . | Age | Age | 465.19 | 24 | 777.87 |
| 6 | . | . | Age | 468.80 | 14 | 760.58 |
| 7* | Age | Age | Age | 409.38 | 33 | 741.27 |
| **8** | **Age** | **.** | **Age** | **412.98** | **23** | **723.55** |

* Assessing the fit of such complex models is not an easy task (Pradel et al. 2003). Given that our model included only age differences in parameters, we pooled data over cohorts and resorted to a parametric bootstrap approach. We simulated 3000 samples using multinomial laws parameterised with the estimates from the best models (RANDMULTINOMIAL

function, SAS for Windows 9.1. SAS institute, Cary, NC, USA). A chi-square statistic was calculated from each sample. As a goodness of fit test, we assessed the number of samples for which the value of the chi-square was superior from the one calculated from original data. In both groups, model 7 fitted satisfactorily the data (*I*: $P = 0.10$; *C*: $P = 0.41$).

**Appendix 2. Summary of the numbers of observations in each state and the observed transitions between states in the survival data (cumulative over several intervals).** This table gives the number of observations in states LAR (larvae), NRA (non-reproductive adult), and FRA (first-time reproductive adult) as well as the number of transitions between LAR and NRA and, NRA and FRA. Transitions from FRA to NRA or LAR, and from NRA to LAR were deemed to be impossible (see text for details). There were few observations as FRA. *I*: Isolation rearing history; *C*: Crowding rearing history.

| State | I | C |
|---|---|---|
| LAR | 260 | 519 |
| NRA | 248 | 244 |
| FRA | 44 | 29 |
| LAR→NRA | 54 | 44 |
| NRA→FRA | 23 | 16 |

**Discussion et Perspectives.** L'approche génétique au service de l'étude des pullulations d'insectes et de l'aide à leur gestion

Le présent travail s'inscrit dans une nouvelle orientation des recherches sur le locuste *L. migratoria*, dont le but est d'apporter une connaissance significative à la fois sur la structure génétique de l'espèce et de ses populations pullulantes et non pullulantes à l'aide d'outils moléculaires, et sur le processus de grégarisation chez les populations pullulantes et non pullulantes à l'aide d'expérimentations en milieu contrôlé. Dans ce contexte, les microsatellites ont satisfait nos attentes, en se révélant informatifs aux niveaux génétiques explorés. L'utilisation de marqueurs microsatellites a permis de réévaluer la structuration géographique mondiale initialement proposée sur la base de l'étude de la morphométrie et d'inférer, au moins sommairement, les facteurs à l'origine de cette structuration. L'information la plus significative apportée par les marqueurs microsatellites, dans le cadre général de l'étude des insectes pullulants, a été la mise en évidence d'un effet homogénéisant des évènements de pullulations sur de vastes échelles géographiques, aussi bien en milieux tempéré que tropical, rendant faibles les chances d'inférer les mouvements des populations pullulantes à partir de l'approche moléculaire. Parallèlement, les expérimentations en élevage ont répondu à nos objectifs en apportant un éclairage évolutif et démographique sur les relations entre grégarisation et pullulation.

Certains aspects du développement, de l'utilisation, et du gain d'information des modèles et des outils de génétique méritent d'être discutés plus avant de par leurs limites ou leur pertinence en biologie des insectes pullulants.

(i) Si les élevages ont nécessité un protocole soigné et se sont avérés très coûteux en temps, la mise au point des marqueurs génétiques ainsi que l'interprétation des données obtenues à ces marqueurs ont été singulièrement délicates. En effet, le succès de l'approche basée sur les marqueurs génétiques chez *L. migratoria* n'a été possible que par des protocoles de biologie moléculaire très lourds et une prise en compte dans les analyses génétiques à la fois des modalités évolutives de variation des séquences d'ADN étudiées, à l'origine d'allèles nuls en forte fréquence, et des caractéristiques populationnelles propres à l'espèce, notamment des grandes tailles de population, une forte dispersion, et une fluctuation dans le temps des tailles de population et des taux de dispersion. Les problèmes rencontrés lors de nos études basées sur des marqueurs moléculaires chez *L. migratoria* semblent concerner un grand nombre d'insectes pullulants, comme en témoigne le faible nombre d'études sur ces espèces en dépit de leur importance socio-économique majeure (voir cependant Bogdanowicz *et al.* 1997 sur *L. dispar* ; Ibrahim *et al.* 2000 sur *S. gregaria*).

(ii) Si l'outil génétique a, en premier lieu, permis de mieux décrire différents aspects de la dynamique du processus de pullulation, il a également fourni quelques informations sur les processus évolutifs qui le sous-tendent.

(iii) Dans la mesure où cette thèse a été initiée suite à des demandes récurrentes d'experts et gestionnaires des pullulations de locustes, elle ne saurait être complète sans discuter de l'intégration des outils et résultats de la génétique des populations aux processus décisionnels concernant le contrôle de ces insectes nuisibles.

(iv) Les populations historiquement peu pullulantes ont révélé une situation énigmatique, faisant peut-être appel à la biologie de la spéciation.

Les quatre points évoqués ci-dessus structureront cette section discussion-perspective.

## 1. LIMITES DES MARQUEURS GENETIQUES POUR L'ETUDE DES PULLULATIONS D'INSECTES

### 1.1. Présence d'allèles nuls aux données microsatellites

Les marqueurs microsatellites des lépidoptères et orthoptères, qui représentent une majorité des insectes pullulants, sont sujets à de fortes prévalences d'allèles nuls. Ce problème semble important dans ces ordres du fait d'une forte instabilité mutationnelle des régions flanquantes (Meglecz *et al.* 2004) et de tailles efficaces de population élevées. Nos travaux sur des jeux de données microsatellites en présence d'allèles nuls, à la fois expérimentaux et par simulations informatiques, ont permis de décrire (i) la distribution des allèles nuls aux niveaux intra-populationnel et inter-populationnels, ainsi que les facteurs évolutifs modifiant ces taux, (ii) les biais générés par la présence d'allèles nuls sur les statistiques couramment utilisées pour décrire la variation génétique à l'intérieur et entre les populations, sur la construction de dendrogrammes, et sur les résultats de l'analyse Bayésienne BAPS de Corander *et al.* (2003), (iii) la performance de différentes méthodes pour corriger les biais d'estimation du niveau de différenciation entre populations, mesuré à l'aide du $F_{ST}$ ou de distances génétiques.

Si ce travail nous a permis d'analyser rigoureusement nos jeux de données microsatellites en présence d'allèles nuls, des problèmes importants restent à évaluer. Tout d'abord, nos simulations ont souligné les limites actuelles lorsque l'on souhaite

corriger un jeu de données génotypiques caractérisé par la présence d'allèles nuls. En effet, les méthodes couramment utilisées pour prendre en compte la présence d'allèles nuls dans les jeux de données attribuent un état unique commun aux différentes populations étudiées. Or, les allèles nuls aux locus microsatellites sont souvent associés à plusieurs états, certains non partagés entre les populations. Une modélisation basée sur la théorie du coalescent (Hudson 1990) représente une voie prometteuse dans la mesure où cette théorie permet de modéliser explicitement et de manière efficace la dynamique évolutive des allèles nuls (*i.e.*, un modèle GSM pour la région répétée, elle-même complètement liée aux séquences flanquantes qui évoluent sous IAM). C'est ce que nous avons fait au cours de cette thèse dans une optique purement descriptive et non inférentielle. Dans une optique inférentielle, il s'agirait d'incorporer *via* la coalescence de tels modèles mutationnels à une méthodologie type MCMC (*Markov Chain Monte Carlo* ; Chib et Greenberg 1995) ou ABC (*Approximate Bayesian Computation* ; Beaumont *et al.* 2002) pour estimer les statistiques non biaisées résumant l'information génétique ainsi que les paramètres populationnels clefs tels que les taux de migration et/ou les temps de divergence entre populations. D'autres caractéristiques mutationnelles des locus microsatellites, telles que les contraintes de taille allèlique (*e.g.*, Feldman *et al.* 1997), pourraient également facilement être intégrées dans cette approche.

Il est important de noter que les estimateurs non biaisés du taux d'allèles nuls et du $F_{ST}$ (algorithme de Dempster *et al.* 1977 et estimateur $F_{ST}^{\{ENA\}}$ présenté dans le Manuscrit 2) font l'hypothèse que l'écart à l'équilibre d'Hardy-Weinberg est exclusivement lié à la présence d'allèles nuls. Ceci sous-entend notamment que les populations sont panmictiques et que les marqueurs microsatellites utilisés ne sont pas sous sélection. Une approche inférentielle basée sur la coalescence permettrait là aussi de modéliser explicitement les relations généalogiques d'un échantillon de gènes d'une population en écart à la panmixie et/ou sous sélection, en intégrant un coefficient de consanguinité et/ou de sélection .

Bien qu'une méthodologie inférentielle basée sur le coalescent semble particulièrement attrayante, elle demeure limitée par le fait que les modèles mutationnel et populationnel spécifiés restent éloignés de la réalité biologique. Par exemple, bien que la plupart des études indiquent que le modèle GSM est le modèle le plus réaliste parmi ceux décrits pour les locus microsatellites (*e.g.*, Cornuet *et al.* 2006), plusieurs

facteurs non pris en compte, tels que la recombinaison et des biais de mutation, peuvent être pertinents pour l'évolution de ces locus (revue dans Estoup *et al.* 2002). De plus, quel que soit le modèle populationnel, le nombre de populations dans le système doit être spécifié alors qu'il est généralement inconnu. L'erreur que ces hypothèses non réalistes introduisent dans l'estimation des paramètres d'intérêt est généralement mal connue, et reste à être évaluée.

### 1.2. Copies nucléaires de séquences mitochondriales

Les marqueurs mitochondriaux ont un fort potentiel pour étudier les relations phylogénétiques et phylogéographiques entre les populations et sous-espèces. Malheureusement, ces marqueurs présentent, au moins chez les locustes, des risques élevés de copies nucléaires (*e.g.*, Gellissen et Michaelis 1987). Il apparaît donc indispensable dans ce groupe taxonomique de s'assurer de l'absence de copies nucléaires lors de la caractérisation de ce type de marqueur. Nous avons envisagé de mettre en évidence la présence ou l'absence de copies nucléaires d'une séquence mitochondriale cible selon la stratégie suivante : dans un premier temps, amplifier la mitochondrie complète (environ 16 kb chez *L. migratoria*) selon le protocole de Hwang *et al.* (2001), dans un second temps, amplifier la séquence cible à partir de cet amplifiat, et enfin comparer les séquences obtenues par cette stratégie avec celles obtenues par un séquençage classique (direct). Amplifier la mitochondrie complète semble difficile mais reste la stratégie la plus sûre puisque des transferts de grands fragments mitochondriaux dans le noyau ont été documentés (*e.g.*, 7.9 kb chez le chat Lopez *et al.* 1994). De plus, les chances de succès de cette stratégie sont accrues chez *L. migratoria* par la possibilité de définir des amorces spécifiques, dans la mesure où la mitochondrie de cette espèce a été complètement séquencée (Flook *et al.* 1995). Nous n'avons malheureusement pas eu le temps de tester cette stratégie, ce qui mériterait d'être fait ultérieurement.

Dans l'espoir d'utiliser plus avant les marqueurs moléculaires dans l'étude des insectes pullulants, il apparaît nécessaire de mieux comprendre les modalités et l'origine de variation de leurs génomes, en particulier des séquences flanquantes des microsatellites dupliquées et/ou fortement variables et des séquences mitochondriales transférées dans le noyau.

### 1.3. Déséquilibre mutation-migration-dérive

De par leurs caractéristiques démographiques, il est peu probable que les insectes pullulants soient à l'équilibre mutation-dérive-migration. Il est donc difficile d'interpréter chez ces organismes les mesures de la variation génétique et de sa distribution en termes de paramètres biologiques tels que les taux de migration et les tailles de population efficaces. Il semblerait pertinent d'évaluer par des simulations informatiques, basées sur la théorie de la coalescence, les conséquences générales des pullulations sur la diversité et la structure génétique des populations. Il s'agirait ici de décrire les signatures génétiques, sur les statistiques résumant la variation génétique à l'intérieur et entre les populations (*e.g.*, hétérozygotie attendue (Nei 1987) et $F_{ST}$ (Weir 1996)) associées à des fluctuations récurrentes des tailles efficaces et des taux de migration entre des populations. Ces signatures génétiques pourront être comparées au cas de populations démographiquement stables. Un bilan des scénarios démographiques d'espèces pullulantes (parasites, mammifères, insectes) serait judicieux afin d'ancrer ces simulations dans un cadre évolutif dépassant celui des locustes.

En attendant de tels développements, il est en pratique possible, moyennant un effort d'échantillonnage et de génotypage considérable, de contourner l'absence de valeurs attendues des statistiques résumant la variation génétique à l'intérieur et entre les populations sous des modèles de déséquilibre mutation-dérive-migration, par l'utilisation d'approches comparatives temporelles ou spatiales. Nous avons adopté cette dernière stratégie en comparant la variation génétique à l'intérieur et entre des populations de *L. migratoria* pullulantes et non pullulantes.

## 2. ECLAIRAGES EVOLUTIFS SUR LES PROCESSUS DE GREGARISATION ET DE PULLULATION CHEZ *L. MIGRATORIA*

### 2.1. Variation génétique des propensions à pulluler et à grégariser

Nous avons recherché s'il existe une adéquation entre la différenciation génétique estimée à des marqueurs sélectivement neutres (*i.e.*, microsatellites) et les différences de la propension à pulluler des populations de *L. migratoria*. La recherche d'une congruence entre ces deux niveaux de variation populationnelle implique avant tout que

les évènements de pullulations chez cette espèce ne sont pas strictement déterminés par des facteurs environnementaux mais également par des traits populationnels sous variation génétique (potentiellement adaptative). Nous n'avons pas mis en évidence de congruence entre la variation aux marqueurs microsatellites et la variation de la propension à pulluler des populations. Cependant, la structure populationnelle de marqueurs neutres tels que les microsatellites ne correspond pas nécessairement à celle de locus sélectionnés. L'adéquation des deux types de structure populationnelle dépend en effet des coefficients de migration et de sélection ainsi que des taux de recombinaison entre les locus des deux catégories. En revanche, l'évolution du trait de pullulation pourrait au moins partiellement refléter l'histoire évolutive neutre des populations *via* des effets des évènements de colonisation et de fondation. Toutefois, les flux de gènes contemporains particulièrement intenses chez *L. migratoria* sont susceptibles de brouiller fortement un tel signal phylogénétique et historique aux marqueurs moléculaires.

Comme les traits de grégarisation sont les meilleurs candidats pour le contrôle du développement ou la persistance des pullulations de locustes, nous avons également évalué plus directement la variation génétique de la propension à grégariser, à l'aide d'une étude d'évolution expérimentale. Nous avons ainsi établi que la variation populationnelle, au moins entre une population malgache et une population française, de la propension à grégariser résulte en partie d'un processus génétique sans doute adaptatif. Si la grégarisation a bien un rôle dans la dynamique de pullulation, on peut s'attendre à ce que la variation génétique de la propension à grégariser résulte en une variation génétique de la propension à pulluler. Une analyse similaire sur le criquet australien *Chortoicetes terminifera* et centrée sur le seul trait phasaire connu chez cette espèce, le comportement d'agrégation et d'activité, est envisagée en collaboration avec Stephen Simpson et Gregory Sword (Université de Sydney, Australie), afin de comparer nos résultats avec ceux d'une autre espèce de locuste.

### 2.2. Avantages adaptatifs à grégariser

Sachant que nous avons à présent montré l'existence d'une variation génétique, *a priori* adaptative, de la propension à grégariser (au moins entre certaines populations de *L. migratoria*), nous pouvons nous interroger sur la valeur adaptative de ce trait. Comme détaillé en introduction, la grégarisation a été interprétée comme une adaptation

à la migration (Ellis 1953 ; traits d'agrégation) ou aux fortes densités de populations, qui entraînent de forts risques de prédation (revue dans Simpson *et al.* 2006; traits d'agrégation et de coloration), de maladie (Wilson *et al.* 2002 ; trait immunitaires), ou de compétition pour la nourriture et les femelles (Simpson *et al.* 2002 ; Seidelmann *et al.* 2005 ; traits physiologiques). L'ensemble de ces hypothèses sur la valeur adaptative du changement de phase ont cependant été exclusivement formulées chez *S. gregaria*. Au cours de cette thèse, nous avons montré que les individus grégaires de *L. migratoria* investissent plus précocement et en plus grande importance dans le premier événement de reproduction. Des études expérimentales et théoriques sur l'évolution des traits d'histoire de vie, en particulier chez les insectes mais aussi sur d'autres groupes, ont mis en évidence que plus les organismes étaient sensibles précocement à la mortalité, plus ils investissaient dans la reproduction (revue dans Reznick 2002 et Fox et Mousseau 1998). Ainsi, nous suggérons que la grégarisation chez *L. migratoria* peut être une adaptation aux fortes pressions de mortalité, liées à la fois aux fortes densités de populations et à une disponibilité en ressources limitée et imprévisible. Ce résultat ainsi que son interprétation sont cependant à prendre avec précaution. En effet, la théorie sur l'évolution des traits d'histoire de vie prédit qu'un investissement plus important dans la reproduction conduit à une fécondité ou une survie ultérieures plus faibles alors que nous n'avons pas détecté dans notre expérience de tels compromis. Ce paradoxe reste pour le moment inexpliqué.

La majorité des espèces de locustes ne sont pas très proches phylogénétiquement, puisqu'elles appartiennent à différentes sous-familles (Fig. 1). Ainsi, les sources de variations et la valeur adaptative du changement de phase discutées chez *L. migratoria* peuvent différer avec les autres espèces de locustes. Il serait judicieux de comparer les résultats obtenus au cours de cette thèse à ceux obtenus pour d'autres espèces de locustes.

## 3. DE LA GESTION DES PULLULATIONS DE *L. MIGRATORIA*

### 3.1. Inférer les sources des pullulations et les routes d'invasion par l'approche moléculaire

Nous avons montré qu'une différenciation génétique aux locus microsatellites est absente ou faible entre les populations d'une même aire d'histoire de pullulations fréquentes et intenses de *L. migratoria*. Ce résultat s'explique surtout par la forte capacité de dispersion de l'espèce qui s'exprime plus particulièrement lors des périodes de pullulation. Cette faible différenciation reflète également les grandes tailles efficaces des populations de *L. migratoria*. En effet, suite à une homogénéisation génétique lors d'une phase de pullulation, les fortes tailles efficaces des populations de *L. migratoria* engendrent un retour à l'équilibre lent vers les valeurs de $F_{ST}$ attendues à l'équilibre. D'autre part, de fortes hétérozygosities aux locus microsatellites sont susceptibles de contraindre fortement l'estimation de la différentiation entre populations calculée à partir de l'estimateur traditionnel $F_{ST}$ (Weir 1996). En effet, les valeurs maximales de $F_{ST}$ pour des populations ne partageant aucun allèle, correspondant à des situations de divergence ancienne sans flux de gènes, ne peut pas dépasser le niveau moyen de l'homozygotie intra-populationnelle (Hedrick 1999). Enfin, et dans une moindre mesure, des études empiriques et numériques ont montré que les situations impliquant de grandes tailles efficaces des populations, associées aux forts taux de mutation et contraintes de taille caractéristiques des modalités évolutives des microsatellites (Bowcock *et al.* 1994; Garza *et al.* 1995), sont susceptibles d'amplifier l'homoplasie de taille, et par là l'homogénéité génétique apparente (Queney *et al.* 2001; Estoup *et al.* 2002). Les fortes capacités de dispersion et les grandes tailles de population sont des traits biologiques communément partagés par les insectes pullulants. Il est donc possible, voire probable, que les patrons de différenciation observés chez *L. migratoria* le soient également chez les autres espèces d'insectes pullulants (voir par exemple Bogdanowicz *et al.* 1997).

Du fait de la faible variation génétique entre les populations de *L. migratoria* dans les zones historiquement pullulantes, il est attendu que les approches génétiques traditionnelles pour inférer les sources et les mouvements de populations, basées sur les *F*-statistiques (Weir 1996) ou les statistiques d'assignation (Rannala et Mountain 1997;

Paetkau *et al.* 2004), soient caractérisées par une faible résolution (*e.g.*, Roeder *et al.* 2001). Une stratégie possible pour palier au manque d'information des données génétiques consisterait à intégrer dans les analyses d'autres types de données telles que des données historiques (*e.g.*, Estoup *et al.* 2001 ; Miller *et al.* 2005), démographiques, et/ou environnementales (Gaggiotti *et al.* 2002). Ainsi, un projet combinant l'analyse de la variation aux marqueurs microsatellites, des densités de populations, et de l'abondance et la distribution des ressources est en cours de développement en collaboration avec Stephen Simpson et Gregory Sword (Université de Sydney, Australie) dans le but de retracer les mouvements des populations du criquet australien *Chortoicetes terminifera* au début d'un évènement de pullulation.

Notons également l'existence d'une méthode récente estimant la proportion d'individus de différentes populations sources ayant fondé une population mélangée (Gaggiotti *et al.* 2002 ; 2004). Cette méthode semble relativement insensible au niveau de différenciation génétique des populations sources. Dans le cas d'une faible différenciation, cette méthode repose sur l'existence d'évènements de fondation impliquant un nombre réduit d'individus susceptible de générer transitoirement des déséquilibres de liaison génétique aux génotypes multilocus. Des compléments d'information non génétique fournis par le modèle, telles que des données démographiques ou de distances géographiques, peuvent également être intégrés dans le processus d'inférence. Dans le contexte spécifique des aires pullulantes de *L. migratoria*, les populations en mélange sont *a priori* très peu différenciées et il est peu probable que des évènements de fondation, impliquant un nombre restreint d'individus, soient associés au processus d'invasion/colonisation. De ce fait, l'information fournie par le déséquilibre de liaison entre les locus, et par là, la puissance de cette méthode, risque d'être limitée. Il serait cependant intéressant d'évaluer la puissance d'une telle méthodologie dans le contexte spécifique de notre modèle d'étude.

### 3.2. Altérer l'expression des traits phasaires impliqués dans la capacité à pulluler

Dans cette thèse, nous avons montré un plus fort investissement dans la reproduction, caractérisé par une reproduction plus précoce et une descendance de meilleure qualité, avec la grégarisation. De par notre protocole expérimental, les effets phase-dépendants mis en évidence sont des effets maternels. Des augmentations

décalées du potentiel reproducteur en conditions de fortes densités pourraient avoir une importance particulière dans la production ou la persistance des pullulations. Il est par ailleurs connu que les traits de migration changent avec la grégarisation, affectant ainsi la dynamique de populations. Dans ce contexte, il semblerait pertinent en gestion des populations de tenter d'altérer l'expression des traits phasaires de reproduction et de comportement. Comme nous avons mis en évidence un contrôle génétique du changement phasaire, nous pouvons imaginer une stratégie basée sur l'interférence ARN pour rendre silencieuse l'expression de ces traits chez *L. migratoria*. Une étape préliminaire pour cela consisterait à comparer l'expression génomique de criquets grégaires et solitaires de populations ayant une forte capacité à pulluller (*e.g.*, Madagascar) et de populations qui ont une capacité à pulluler moindre (*e.g.*, France). Une telle étude a été envisagée, mais non réalisée par manque de temps et de moyen, au cours de cette thèse. Dans ce but, l'ARN d'individus français et malgache de phase grégaire et solitaire induite au laboratoire, selon le protocole d'élevage présenté dans le chapitre III, a été conservé.

## 4. QUID DES POPULATIONS NON PULLULANTES ?

Dans les zones situées en limite d'aire de répartition de l'espèce et historiquement non-pullulantes, un nombre non négligeable de populations de *L. migratoria* ont montré un niveau de différenciation génétique considérable entre eux et par rapport aux grands ensembles populationnels géographiquement proches (*e.g.*, les populations de la forme de « Palavas », l'Espiguette et les Aresquiers, ainsi que la population du Nord de la France). Ces populations montrent des diversités génétiques inférieures et des niveaux de différenciation supérieures aux populations connectées de zones pullulantes (*e.g.*, Madagascar) mais également non-pullulantes (*e.g.*, côte ouest méditerranéenne). Ceci s'explique probablement par des processus neutres, tels que les effets d'isolement géographique, de fondation et de dérive, mais aussi potentiellement par des processus adaptatifs, rendus possibles par l'absence de flux de gènes avec les populations principales. Nous avons par exemple montré que la population d'Europe du Nord de la France n'échangeait pas ou peu de gènes avec les autres populations, probablement du fait de son isolement géographique. Les habitats de cette aire géographique, située en limite d'aire de répartition de l'espèce, ont potentiellement une capacité biotique faible,

ce qui impliquerait des tailles efficaces des populations concernées relativement faibles. La population du Nord de la France est une population récente, découverte en 1996, probablement fondée par une ou des propagules issues d'environnements plus favorables géographiquement proches, tels que les littoraux océaniques. En accord avec cette hypothèse, la population du Nord de la France est génétiquement rattachée à la population littorale de Lacanau. Enfin, il est envisageable que certaines de ces populations se soient adaptées aux conditions extrêmes et locales de leurs aires d'habitat.

Les populations de l'Espiguette et des Aresquiers, qui limitent la zone de Palavas, sont un cas énigmatique quant aux causes de leur isolement génétique. En effet, ces populations génétiquement isolées sont localisées dans un écosystème de lagunes qui offre des aires humides particulièrement favorables à *L. migratoria* (Remaudière 1940a; 1940b), le long de la côte Méditerranéenne où de nombreuses autres populations sont interconnectées. Cette situation rend peu probables les hypothèses de dérive génétique associée à une faible capacité biotique du milieu et d'absence de flux de gènes liée à un isolement géographique. Une alternative historique a été suggéré par Remaudière (1940a) : des individus migrants venant de la côte ouest de la Mer Noire et appartenant à la sous-espèce *L. m. migratoria* auraient fondé les populations de Palavas durant la dernière pullulation enregistrée dans cette aire au XIV$^{ème}$ siècle. Ce scénario pourrait expliquer les forts niveaux de différenciation ainsi que les plus faibles diversités des populations de Palavas mais ne résout pas l'absence de flux de gènes avec les populations environnantes de *L. m. cinerascens* mis en évidence par l'analyse Bayésienne 2MOD (Ciofi *et al.* 1999). Des mécanismes d'isolement pré ou post-reproducteurs pourraient alors être impliqués dans cette surprenante absence de flux de gènes. Les adultes de « Palavas » étant connus pour leur grande taille (Remaudière 1940a), il est concevable qu'un isolement pré-reproducteur, par de l'homogamie de taille ait évolué, rendant la migration non efficace entre les populations de type « Palavas » et *L. m. cinerascens*. Notons toutefois que cette hypothèse n'explique pas la forte différenciation génétique observée entre les deux populations génotypées de la forme de Palavas, situées seulement à 50 km l'une de l'autre. Des mesures morphologiques et génétiques d'échantillons de populations couplées à une caractérisation du milieu (*e.g.*, disponibilité et distribution des ressources, température, humidité) le long de transects sur le littoral méditerranéen, incluant les lagunes de Palavas, permettraient de préciser le statut d'isolement des populations des lagunes de

Palavas et les facteurs évolutifs ou écologiques qui le façonne. Des expérimentations en laboratoire et de terrain sur le comportement d'accouplement seraient nécessaires pour tester l'hypothèse d'homogamie par la taille.

# Bibliographie

Albrecht FO, Verdier M, Blackith RE (1959) Maternal control of ovariole number in the progeny of the migratory locust. *Nature* **184**, 103-104.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Applebaum SW, Heifetz Y (1999) Density-dependent physiological phase in insects. *Annual Review of Entomology* **44**, 317-341.

Avise JC (2000) *Phylogeography: the History and Formation of Species*, Harward University Press, Cambridge.

Babah MAO, Sword GA (2004) Linking locust gregarization to local resource distribution patterns across a large spatial scale. *Environmental Entomology* **33**, 1577-1583.

Bailey NW, Gwynne DT, Ritchie MG (2005b) Are solitary and gregarious Mormon crickets (*Anabrus simplex*, Orthoptera, Tettigoniidae) genetically distinct? *Heredity* **95**, 166-173.

Bailey NW, Hockham LR, Graves JA, Ritchie MG (2005a) Development and characterization of microsatellite loci in Mormon crickets (*Anabrus simplex*, Orthoptera,Tettigoniidae). *Molecular Ecology Notes* **5**, 613-615.

Baltensweiler W (1984) The role of environment and reproduction in the population dynamics of the larch bud moth, *Zeiraphera diniana* Gn. (Lepidoptera, Tortricidae). *Advances in Invertebrate Reproduction* **3**, 291-302.

Beaumont MA, Zhang WJ, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035.

Beckerman A, Benton TG, Ranta E, Kaitala V, Lundberg P (2002) Population dynamic consequences of delayed life-history effects. *Trends in Ecology and Evolution* **17**, 263-269.

Benfekih L, Chara B, Doumandji-Mitiche B (2002) Influence of anthropogenic impact on the habitats and swarming risks of *Dociostaurus maroccanus* and *Locusta migratoria* (Orthoptera, Acrididae) in the Algerian Sahara and the semiarid zone. *Journal of Orthoptera Research* **11**, 243-250.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**, 314-321.

Bernardini C, Ketmaier V (2002) Isolation and characterization of five microsatellite loci in Dolichopoda cave crickets (Orthoptera Rhaphidophoridae). *Molecular Ecology Notes* **2**, 501-503.

Berryman AA (1987) The theory and classification of outbreaks. *in P. Barbosa and J.C. Schultz, eds. Insect outbreaks. Academic Press, New York*, 3-30.

Berthier K, Charbonnel N, Galan M, Chaval Y, Cosson JF (2006) Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. *Molecular Ecology* **in press**.

Betts E (1961) Outbreaks of the African migratory locust (*Locusta migratoria migratorioides* R. & F.) since 1871. *Anti-Locust Memoir* **6**, 1-25.

Billings RF, Berisford CW, Salom SM, Payne TL (1995) Application of semiochemicals in the management of southern pine beetle infestations: current status of research. *in Application of Semiochemicals for Management of Bark Beetle Infestations-Proceedings of an Informal Conference. eds Salom S.M.*

*and Hobson K.R. December 12-16, 1994, Dallas, TX. USDA, Forest Service, Intermountain Res. Stat., Gen. Tech. Rep. No. 318.*, 30-38.

Black WC, Krafsur ES (1986) Seasonal breeding structure in house fly, *Musca domestica* L., populations. *Heredity* **56**, 289-298.

Blackith RE, Albrecht FO (1979) Locust plagues: the interplay of endogenous and exogenous control. *Acrida* **8**, 83-94.

Bogdanowicz SM, Mastro VC, Prasher DC, Harrison RG (1997) Microsatellite DNA variation among asian and north american gypsy moths (Lepidoptera: Lymantriidae). *Annals of the Entomological Society of America* **90**, 768-775.

Borden JH (1989) Semiochemicals and bark beetle populations: Expoitation of natural phenomena by pest management stategies. *Holarctic Ecology* **12**, 501-510.

Bouaïchi A, Roessingh P, Simpson SJ (1995) An analysis of the behavioural effects of crowding and re-isolation on solitary-reared adult desert locusts (*Shistocerca gregaria*) and their offspring. *Physiological Entomology* **20**, 199-208.

Bowcock AM, Ruiz-Linares A, Tomfohrde J*, et al.* (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-457.

Breuer M, Hoste B, De Loof A (2003) The endocrine control of phase transition: some new aspects. *Physiological Entomology* **28**, 3-10.

Brown HD (1986) New locust problem. *Antenna* **10**, 11-13.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Chen Y-L (1991) The migratory locust, *Locusta migratoria*, and its asiatic subspecies. *Ortopterists' Society & Lyman Entomological Museum, Québec*.

Chib S, Greenberg E (1995) Understanding the Metropolis-Hastings algorithm. *American Statistician* **49**, 327-335.

Chitty D (1967) The natural selection of self-regulatory behaviour in animal populations. *Proceedings of the Ecological Society of Australia* **2**, 51-78.

Ciofi C, Beaumont MA, Swingland IR, Bruford MW (1999) Genetic divergence and units for conservation in the Komodo dragon *Varanus komodensis*. *Proceedings of Royal Society of London serie B* **266**, 2269-2274.

Collett R (1912) *Norges Pattedyr*, Aschehoug, Kristiania.

COPR (1982) *The Locust and Grasshopper Agricultural Manual*, Centre for Overseas Pest Research, London.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

Cornuet JM, Beaumont MA, Estoup A, Solignac M (2006) Inference on microsatellite mutation processes in the invasive mite, Varroa destructor, using reversible jump Markov Chain Monte Carlo. *Theoretical Population Biology* **69**, 129-144.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* **93**, 504-509.

Davey JT (1956) The seasonal migrations and dynamics of populations of the African migratory Locust (*Locusta migratoria migratorioides* Reiche and Fairmaire) in the outbreak area (Orth.). *Bulletin de la Société Entomologique de France* **61**, 18-24.

David P, Samadi S (2000) *La Théorie de l'Evolution. Une logique pour la biologie.*, Editions Flammarion, Paris.

De Loof A, Claeys I, Simonet G*, et al.* (2006) Molecular markers of phase transition in locusts. *Insect Science* **sous presse**.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data *via* the EM algorithm. *Journal of the Royal Statistical Society Series B* **39**, 1-38.

Dempster JP (1963) The population dynamics of grasshoppers and locusts. *Biological Reviews* **38**, 490-529.

Despland E, Collett M, Simpson SJ (2000) Small-scale processes in desert locust swarm formation: how vegetation patterns influence gregarization. *Oikos* **88**, 652-662.

Dirsh VM (1953) Morphometrical studies on phases of the desert locust. *Anti Locust Bulletin* **16**, 1-34.

Duranton JF, Launois M, Launois-Luong MH, Lecoq M (1979) Les voies privilégiées de déplacement du Criquet migrateur malgache en phase solitaire. *Bulletin d'Ecologie* **10**, 107-123.

Elkinton JS, Dwyer G, Sharov A (1995) Modelling the epizootiology of gypsy-moth nuclear polyhedrosis-virus. *Computers and Electronics in Agriculture* **13**, 91-102.

Ellis PE (1953) The gregarious behaviour of marching *Locusta migratoria migratorioides* (R. & F.) hoppers. *The journal of experimental biology* **30(2)**, 214-234.

Elton CS (1924) Periodic fluctuations in the numbers of animals: their causes and effects. *British Journal of Experimental Biology* **2**, 119-163.

Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. *in Advances in Molecular Ecology. eds. Carvalho G. (Amsterdam NATO press)*, 55-86. Amsterdam NATO Press.

Estoup A, Garnery L, Solignac M, Cornuet JM (1995) Microsatellite variation in honey bee (*Apis mellifera*) populations - Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* **140**, 679-695.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consquences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.

Estoup A, Wilson IJ, Sullivan C, Cornuet JM, Moritz C (2001) Inferring Population History from Microsatellite and Enzyme Data in Serially Introduced Cane Toads, *Bufo marinus*. *Genetics*.

Farrow RA (1975) The african migratory locust in its main outbreak area of the middle Niger: quantitative studies of solitary populations in relation to environmental factors. *Locusta 11: 198p*.

Farrow RA (1979) Causes of recent changes in the distribution and abundance of the migratory locust (*Locusta migratoria* L.) in Australia in relation to plagues. *CSIRO Division of Entomology Commonwealth Scientific and Industrial Organization, Canberra* **9**.

Farrow RA (1979) Causes of recent changes in the distribution and abundance of the migratory locust (*Locusta migratoria* L.) in Australia in relation to plagues. *CSIRO Division of Entomology Commonwealth Scientific and Industrial Organization, Canberra* **9**.

Farrow RA (1987) Effects of changing land use on outbreaks of tropical migratory locust, *Locusta migratoria migratorioides* (R. & F.). *Insect Scientific Application* **8**, 969-975.

Farrow RA (1990) Flight and migration in Acridoids. *in Biology of Grasshoppers. eds. Chapman R.F. and Joern A.*, 227-314. John Wiley & Sons.

Farrow RA, Colless DH (1980) Analysis of the Interrelationships of geographical races of *Locusta migratoria* (Linnaeus) (Orthoptera : Acrididae) by numerical taxonomy, with special reference to sub-speciation in the tropics and affinities of the australian race. *Acrida* **9**, 77-99.

Flook PK, Rowell CHF, Gellissen G (1995) The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *Journal of Molecular Evolution* **41**, 928-941.

Fox CW, Mousseau TA (1998) Maternal effects as adaptations for transgenerational phenotypic plasticity in insects. *in T.A. Mousseau and C.W. Fox, Maternal effects as adaptations. Oxford University Press, New-York*, 159-177.

Franc A, Rabesisoa FL, Luong-Skovmand MH, Lecoq M (2005) Phase polymorphism in the red locust, *Nomadacris septemfasciata* Serville (Orthoptera: Cyrtacanthacridinae) in Madagascar. *International Journal of Tropical Insect Science* **25**, 182-189.

Frankham R (1997) Do island populations have less genetic variation than mainland populations? *Heredity* **78**, 311-327.

Gaggiotti OE, Brooks SP, Amos W, Harwood J (2004) Combining demographic, environmental and genetic data to test hypotheses about colonization events in metapopulations. *Molecular Ecology* **13**, 811-825.

Gaggiotti OE, Jones F, Lee WM*, et al.* (2002) Patterns of colonization in a metapopulation of grey seals. *Nature* **416**, 424-427.

Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6**, 861-868.

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular and Biological Evolution* **12**, 594-603.

Gellissen G, Michaelis G (1987) Gene transfer mitochondria to nucleus. *Annals of New York Academy of Sciences* **503**, 391-401.

Ginzburg LR (1998) Inertial growth: population dynamics based on maternal effects. *in Maternal effects as Adaptations. eds Mousseau T.A. and Fox C.W.*, 42-53. Oxford University Press, New-York.

Ginzburg LR, Taneyhill DE (1994) Population cycles of forest Lepidoptera: a maternal effect hypothesis. *Journal of Animal Ecology* **1994**, 79-92.

Glize E (1996) L'invasion des criquets dans les Landes. *Bulletin de la Société de Borda* **121**, 173-194.

Hägele BF, Simpson SJ (2000) The influence of mechanical, visual and contact chemical stimulation on the behavioural phase state of solitarious desert locusts (*Schistocerca gregaria*). *Journal of Insect Physiology* **46**, 1295-1301.

Hakomori T, Tanaka S (1992) Genetic control of diapause and other developmental traits in Japanese strains of the migratory locust, *Locusta migratoria L.* : univoltine vs. bivoltine. *Japanese Journal of Entomology* **60**, 319-328.

Hancock JM, Armstrong JS (1994) SIMPLE34: an improved and enhanced implementation for VAX and Sun computers of the SIMPLE algorithm for analysis of clustered repetitive motifs in nucleotide sequences. *Computer Applications in the Biosciences* **10**, 67-70.

Hassanali A, Bashir MO (1999) Insights for the management of different locust species from new findings on the chemical ecology of the desert locust. *Insect Science and its Application* **19**, 369-376.

Hassanali A, Njagi PGN, Bashir MO (2005) Chemical ecology of locusts and related acridids. *Annual Review of Entomology* **50**, 223-245.

Hedrick PW (1999) perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313-318.

Heifetz Y, Applebaum SW, Popov GB (1994) Phase characteristics of the Israëli population of the migratory locust, *Locusta migratoria* (L.) (Orthoptera: Acrididae). *Journal of Orthoptera Research* **2**, 15-20.

Hockham LR, Graves JA, Ritchie MG (1999) Isolation and characterization of microsatellite loci in the bushcricket *Ephippiger ephippiger* (Orthoptera : Tettigoniidae). *Molecular Ecology* **9**, 905-906.

Hodges JD, Elam WW, Watson WR, Nebeker TE (1979) Oleoresin characteristics and susceptibility of four southern pines to southern pine beetle (Coleoptera: Scolytidae) attacks. *Canadian Entomologist* **111**, 889-896.

Hoste B, Simpson SJ, Tanaka S, De Loof A, Breuer M (2002) A comparison of phase-related shifts in behavior and morphometrics of an albino strain, deficient in [His7]-corazonin, and a normally colored *Locusta migratoria* strain. *Journal of Insect Physiology* **48**, 791-801.

Hoy MA (1994) *Insect Molecular Genetics, Second Edition*, Academic Press/Elsevier, San Diego, CA.

Hudson RR (1990) Gene genealogies and the coalescent process in Oxford Surveys in Evolutionary Biology. *eds. Futuyama, D.J. & Antonovics, J. (Oxford University Press)*, 1-44.

Hunter DM (2004) Advances in the control of locusts (Orthoptera: Acrididae) in eastern Australia: from crop protection to preventive control. *Australian Journal of Entomology* **43**, 293-303.

Hunter DM, Zhang L (2005) Laboratory and field trials of Green Guard (*Metarhizium anisopliae var. acridum*) (Deuteromycotina: Hyphomycetes) against the oriental migratory locust (*Locusta migratoria manilensis*) (Orthoptera: Acrididae) in China. *Journal of Orthoptera Research* **14**, 31-34.

Ibrahim KM, Sourrouille P, Hewitt GM (2000) Are recession populations of the desert locust (*Schistocerca gregaria*) remnants of past swarms? *Molecular Ecology* **9**, 783-791.

Islam MS, Roessingh P, Simpson SJ, McCaffery AR (1994b) Parental effects on the behaviour and colouration of nymphs of the desert locust *Schistocerca gregaria*. *Journal of Insect Physiology* **40**, 173-181.

Islam MS, Roessingh P, Sinervo B, McCaffery AR (1994a) Effects of population density experienced by parents during mating and oviposition on the phase of hatchlings desert locusts, *Schistocerca gregaria*. *Proceedings of Royal Society of London serie B* **257**, 93-98.

Itô Y (1968) Are insect populations periodically fluctuating? *Kagaku* **38**, 39-45.

Jannone G (1947) Italy: numerous outbreaks of *Locusta migratoria* L. phase gregaria (typical) and phases of transition in the province of Naples. *International Bulletin of Plant Protection* **10**, 218-219.

Joly P (1951) Déterminisme endocrine de la pigmentation chez *Locusta migratoria* L. *Comptes Rendus des Séances de la Société de Biologie* **145**, 1362-1364.

Kang L, Chen X, Zhou Y*, et al.* (2004) The analysis of large-scale gene expresion correlated to the phase changes of the migratory locust. *Proceedings of the National Academy of Sciences USA* **101**, 17611-17615.

Karabag T (1958) The Orthoptera Fauna of Turkey. A synonymic and distributional Catalogue of Turkish Orthoptera. *Ankara University Sciences Faculty, Ankara, Turkey.*

Kijas JMH, Fowler JCS, Garbet CA, Thomas MR (1994) Enrichment of microsatellites from citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particules. *Biotechniques* **16**, 657-662.

King TM, Hanotte O, Burke T, Wallis GP (1998) Characterization of four microsatellite loci in tree weta (Orthoptera : Stenopelmatidae): their potential usefulness for the study of Hemideina. *Molecular Ecology* **7**, 663-664.

Lauga J (1977) Le problème de la mesure de la phase chez les acridiens migrateurs : historique et définition d'échelles phasaires chez *Locusta migratoria* L. (Insecte, orthoptère). *Archives de Zoologie Expérimentale et Générale* **118**, 247-272.

Lauga J, Hatté M (1978) L'activité grégarisante du sable de ponte chez *Locusta migratoria* L. : action sur le comportement et la reproduction des individus. *Annales des Sciences naturelles* **20**, 37-52.

Launois M (1974) *Influence du facteur pluviométrique sur l'évolution saisonnière du criquet migrateur Locusta migratoria capito (Saussure) en phase solitaire et sur sa grégarisation à Madagascar*, Ministère de la Coopération, Paris.

Launois M (1978) *Manuel Pratique d'Identification des Principaux Acridiens du Sahel*, GERDAT, Ministère de la Coopération.

Leblois R, Rousset F, Estoup A (2004) Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166**, 1081-1092.

Lecoq M (1975) Les déplacements par vol du criquet migrateur malgache en phase solitaire : leur importance sur la dynamique des populations et la grégarisation. *Thèse de Doctorat d'Etat ès Sciences (Université Paris XI, Orsay). Ministère de la coopération (Paris) : 272 p.*

Lecoq M (2001) Recent progress in Desert and Migratory Locust management in Africa. Are preventive actions possible? *Journal of Orthoptera Research* **10**, 277-291.

Lecoq M, Sukirno (1999) Drought and exceptional outbreak of the oriental migratory locust, *Locusta migratoria manilensis* (meyen 1835) in Indonesia (Orthoptera: Acrididae). *Journal of Orthoptera Research* **8**, 153-161.

Lomer CJ, Bateman RP, Johnson DL, Langewald J, Thomas M (2001) Biological control of locusts and grasshoppers. *Annual Review of Entomology* **46**, 667-702.

Lopez JV, Yuhki N, Masuda R, Modi W, OBrien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial-dna to the nuclear genome of the domestic cat. *Journal of Molecular Evolution* **39**, 174-190.

Lotka AJ (1925) *Elements of Physical Biology*, Williams & Wilkins, Baltimore.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**, 981-994.

Ma S-C (1958) The population dynamics of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China. *Acta Entomologica Sinica* **8**, 1-40.

Mason RR (1987) Nonoutbreak species of forest Lepidoptera. *in P. Barbosa and J.C. Schultz, eds. Insect outbreaks. Academic Press, New York*, 31-58.

Mattson WJ, Haack RA (1987) The role of drought in outbreaks of plant-eating insects. *BioScience* **37**, 110-118.

McCaffery AR, Simpson SJ, Islam MS, Roessingh P (1998) A gregarizing factor present in the egg poad foam of the desert locust *Schistocerca gregaria*. *The Journal of Experimental Biology* **201**, 347-363.

Meglecz E, Petenian F, Danchin E*, et al.* (2004) High similarity between flanking regions of different microsatellites detected within each of two species of lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* **13**, 1693-1700.

Meyen FJF (1835) *Reise um die Erde.*, Eichhoff C.W., Berlin.

Miller N, Estoup A, Toepfer S*, et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science* **310**, 992-992.

Mitter C, Schneider JC (1987) Genetic change and insect outbreaks. *in P. Barbosa and J.C. Schultz, eds. Insect outbreaks. Academic Press, New York*, 502-532.

Motro U, Thomson G (1982) On heterozygosity and the effective size of populations subject to size changes. *Evolution* **36**, 1059-1066.

Nei M (1987) *Molecular Evolutionary Genetics*, Columbia University Press.

Nicholson AJ (1933) The balance of animal populations. *Journal of Animal Ecology* **Supplement 2**, 132-178.

Nicholson AJ (1954) An outline of the dynamics of animal populations. *Australian Journal of Zoology* **2**, 9-65.

Nolte DJ (1973) The relationships between melanin and locustol of locusts. *Acrida* **2**, 61-67.

Norris NJ (1950) Reproduction in the African Migratory locust (*Locusta migratoria migratorioides* R. & F.) in relation to density and phase. *Anti-Locust Bulletin* **6**, 1-50.

Paetkau A, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55-65.

Pener MP (1991) Locust phase polymorphism and its endocrine relations. *Advances in Insect Physiology* **23**, 1-79.

Pener MP, Yerushalmi Y (1998) The physiology of locust phase polymorphism: an update. *Journal of Insect Physiology* **44**, 365-377.

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the european rabbit (*Oryctolagus cuniculus*). *Molecular Biology and Evolution* **18**, 2169-2178.

Randriatmanantsoa M (1998) Manuel sur la lutte antiacridienne. *Direction de la protection des végétaux et Deutsche Gesellschaft für Technische Zusammenarbeit GmbH, Antananarivo*.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* **94**, 9197-9221.

Remaudière G (1940a) Contribution à l'étude des *Locusta migratoria migratoria* L. ph. solitaria de la région de Palavas (Hérault) 1ère partie. *Revue de Pathologie Végétale et d'Entomologie Agricole de France* **17**, 147-163.

Remaudière G (1940b) Contribution à l'étude des *Locusta migratoria migratoria* L. ph. solitaria de la région de Palavas (Hérault) 2ère partie. *Revue de Pathologie Végétale et d'Entomologie Agricole de France* **17**, 220-235.

Remaudière G (1947) Sur l'existence en France d'une nouvelle sous-espèce de *Locusta migratoria* L. *Comptes Rendus des Séances de l'Académie des Sciences* **225**, 1025-1026.

Reznick D, Bryant MJ, Bashey F (2002) *r*- and *K*-selection revisited: the role of population regulation in life-history evolution. *Ecology* **83**, 1509-1520.

Roeder AM, Marshall RK, Mitchelson AJ*, et al.* (2001) Gene flow on the ice: genetic differentiation among Adélie penguin colonies around Antarctica. *Molecular Ecology* **10**, 1645-1656.

Roessingh P, Bouaïchi A, Simpson SJ (1998) Effects of sensory stimuli on the behavioural phase state of the desert locust, Schistocerca gregaria. *Journal of Insect Physiology* **44**, 883-893.

Roffey J, Popov GB (1968) Environmental and behavioural processes in a desert locust outbreak. *Nature* **219**, 446-450.

Rossiter MC (1992) The impact of resource variation on population quality in herbivorous insects: a critical component of population dynamics. *in Resource Distribution and Animal-Plant Interactions. eds. Hunter M.D. , Ohgushi T. , and Price P.W.*, 13-42. Acadamic Press, San diego, CA.

Rossiter MC (1994) Maternal effects hypothesis of herbivore outbreak. *BioScience* **44**, 752-762.

Rossiter MC (1996) The incidence and consequences of inherited environmental effects. *Annual Review of Ecology and Systematics* **27**, 451-476.

Rossiter MC (1998) The role of environmental variation in parental effects expression. *in Maternal effects as Adaptations. eds Mousseau T.A. and Fox C.W.*, 112-134. Oxford University Press, New-York.

Roubaud E (1947) Au sujet du problème acridien en France, différences radicales évolutives entre les peuplements du Criquet migrateur (*Locusta migratoria* L.) des Landes du Sud-Ouest et ceux de la région méditerranéenne. *Comptes Rendus des Séances de l'Académie des Sciences* **225**, 909-911.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstruction of phylogenetic trees. *Molecular and Biological Evolution* **4**, 406-425.

Seidelmann K, Warnstorff K, Ferenz H-J (2005) Phenylacetonitrile is a male specific repellent in gregarious desert locusts, *Schistocerca gregaria*. *Chemoecology* **15**, 37-43.

Shinde D, Lai YL, Sun FZ, Arnheim N (2003) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites. *Nucleic Acid Research* **31**, 974-980.

Simpson SJ, Despland E, Hägele BF, Dodgson T (2001) Gregarious behavior in desert locusts is evoked by touching their back legs. *Proceedings of the National Academy of Sciences USA* **98**, 3895-3897.

Simpson SJ, McCaffery AR, Hägele BF (1999) A behavioural analysis of phase change in the desert locust. *Biological reviews* **74**, 461-480.

Simpson SJ, Raubenheimer D, Behmer ST, Whitworth A, Wright GA (2002) A comparison of nutritional regulation in solitarious- and gregarious-phase nymphs of the desert locust, *Schistocerca gregaria*. *Journal of Experimental Biology* **205**, 121-129.

Simpson SJ, Sword GA, De Loof A (2006) Advances, controversies and consensus in locust phase polyphenism research. *Journal of Orthoptera Research* **sous presse**.

Stenseth NC, Ims RA (1993) Population dynamics of lemmings: temporal and spatial variation - an introduction. *in The Biology of Lemmings. eds. Stenseth NC.*, 61-96. Academic Press, London.

Stenseth NC, Saitoh T, Yoccoz NG (1998) Frontiers in population ecology of microtine rodents: a pluralistic approach to the study of population ecology. *Research in Population Ecology* **40**, 5-20.

Sword GA (2003) To be or not to be a locust? A comparative analysis of behavioral phase change in nymphs of *Schistocerca americana* and *S. gregaria*. *Journal of Insect Physiology* **49**, 709-717.

Sword GA, Lorch PD, Gwynne DT (2005) Migratory bands give crickets protection. **433**, 703-703.

Sword GA, Simpson SJ (2000) Is there an intraspecific role for density-dependent colour change in the desert locust? *Animal Behaviour* **59**, 861-870.

Têtefort JP, Wintrebert D (1966) Effets édaphiques et biotiques de l'inversion pluviométrique des saisons, important facteur de pullulation des acridiens migrateurs dans le sud-ouest malgache. *Entomophaga* **2**, 305-310.

Thomas MB (1999) Ecological approaches and the development of "truly integrated" pest management. *Proceedings of the National Academy of Sciences USA* **96**, 5944-5951.

Turchin P (2003) *Complex Population Dynamics: a Theoretical/Empirical Synthesis*, Princeton University Press, Princeton, NJ.

Uvarov BP (1921) A revision of the genus *Locusta* L, with a new theory as to the periodicity and migrations of locusts. *Bulletin of Entomological Research* **12**, 135-163.

Uvarov BP (1966) *Grasshoppers and Locusts* **Volume 1**, Cambridge University Press.

Uvarov BP, Hamilton AG (1936) Phase variation and rate of development in the algerian race of the migratory locust (Locusta migratoria, L.). *Bulletin of Entomological Research* **27(1)**, 87-90.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

van't Hof AE, Zwaan BJ, Saccheri IJ*, et al.* (2005) Characterization of 28 microsatellite loci for the butterfly *Bicyclus anynana*. *Molecular Ecology Notes* **5**, 169-172.

Vaughan HE, Heslop-Harrison JS, Hewitt GM (1999) The localization of mitochondrial sequences to chromosomal DNA in orthopterans. *Genome* **42**, 874-880.

Volterra V (1926) Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Memorie della R. Accademia nazionale dei Lincei* **2**, 31-113.

Waloff Z (1962) Local outbreaks of the African migratory locust in the Red Sea area. *FAO Plant Protection Bulletin* **10**, 61-63.

Waloff ZV (1940) The distribution and migrations of *Locusta* in europe. *Bulletin of Entomological Research* **31**, 211-246.

Weir BS (1996) *Genetic Data Analysis II*, Sinauer, Sunderland, MA.

Wilson K, Thomas M, Blanford S*, et al.* (2002) Coping with crowds: Density-dependent disease resistance in desert locusts. *Proceedings of the National Academy of Sciences USA* **99**, 5471-5475.

Wintrebert D (1970) Identité, écologie et comportement du criquet migrateur dans le sud-ouest Malgache. *Annales de la Société Entomologique de France* **6**, 35-152.

Wood DL (1972) Selection and colonization of ponderosa pine by bark beetles. *in Insect/Plant Relationships. eds. van Emden H.F.*, 101-117. Blackwell Scientific Publications, Oxford.

Yerushalmi Y, Tauber E, Pener MP (2001) Phase polymorphism in *Locusta migratoria*: the relative effects of geographical strains and albinism on morphometrics. *Physiological Entomology* **26**, 95-105.

Zera AJ, Harshman LG (2001) The physiology of life history trade-offs in animals. *Annual Review of Ecology and Systematics* **2001**, 95-126.

Zhang DX, Hewitt GM (1996) Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Shistocerca gregaria*: some implications for population studies. *Molecular Ecology* **5**, 295-300.

Zhang DX, Yan LN, Ji YJ*, et al.* (2003) Isolation, characterization and cross-species amplification of eight microsatellite DNA loci in the migratory locust (*Locusta migratoria*). *Molecular Ecology Notes* **3**, 483-486.

Zhang Z, Li D (1999) A possible relationship between outbreaks of the oriental migratory locust (*Locusta migratoria manilensis* Meyen) in China and the El Nino episodes. *Ecological Research* **14**, 267-270.

Zolotarevsky BN (1929) Le criquet migrateur (*Locusta migratoria capito* Sauss.) à Madagascar. *Annales des Epiphyties* **15**, 185-240.

## Résumé

Certains insectes ravageurs des cultures et des forêts présentent des fluctuations extrêmes et soudaines des densités de leurs populations. Les études évolutives, basées sur des marqueurs moléculaires neutres et/ou des traits populationnels densité-dépendants, sont potentiellement informatives sur (i) la dynamique et la structure des populations pullulantes et (ii) les déterminants populationnels et environnementaux des pullulations de ces insectes nuisibles. De telles études sont cependant rares chez les insectes pullulants, notamment en raison de modalités évolutives de variation des séquences microsatellites à l'origine d'allèles nuls en forte fréquence.

Dans cette thèse, nous avons quantifié par simulations informatiques la sur-estimation du niveau de différentiation génétique entre les populations et la sous-estimation du niveau de diversité génétique au sein des populations causées par la présence d'allèles nuls aux marqueurs microsatellites. Nous proposons un estimateur du $F_{ST}$ non biaisé, basé uniquement sur les états visibles dont les fréquences génotypiques ont été modifiées selon la fréquence d'allèles nuls.

Cette connaissance sur les allèles nuls aux marqueurs microsatellites a permis une analyse pertinente de la variation génétique à ces marqueurs des populations du criquet migrateur, *Locusta migratoria*, un insecte pullulant qui présente un changement d'une phase solitaire, inactive et dispersée, à une phase grégaire, très mobile et agrégative, lors des évènements de pullulations. La structuration génétique mondiale de cette espèce cosmopolite s'est révélée largement incongruente avec la classification taxonomique actuelle réalisée sur la base de critères morphométriques et comprenant onze sous-espèces. Des facteurs géographiques, écologiques, et historiques structurent la variation génétique de l'espèce à l'échelle de l'ensemble de son aire de répartition, qui en revanche coïncide peu avec le statut pullulant des populations échantillonnées. A une échelle plus locale, nous avons mis en évidence un effet homogénéisant des évènements de pullulations sur de vastes échelles géographiques.

Parallèlement, nous avons réalisé des expérimentations en élevage afin d'apporter un éclairage évolutif et démographique sur les relations entre la capacité à grégariser et la propension à pulluler chez *L. migratoria*. La variation populationnelle, au moins entre une population malgache historiquement pullulante et une population française historiquement non pullulante, de la propension à grégariser résulte en partie d'un processus génétique sans doute adaptatif. Le potentiel reproducteur d'une population malgache historiquement pullulante augmente en phase grégaire, à la fois par une reproduction plus précoce et une meilleur qualité de la descendance, ce qui suggère un rôle du processus de grégarisation dans la croissance numérique des populations durant le développement et/ou le maintien des pullulations.

En termes de gestion des populations pullulantes de *L. migratoria*, ce travail prévient des chances faibles d'inférer les sources des pullulations et les routes d'invasion à partir de l'approche moléculaire du fait de la structuration génétique absente ou minime dans les aires pullulantes. Cependant, il ouvre la possibilité d'une stratégie de gestion basée sur l'altération de l'expression des gènes impliqués dans la capacité à grégariser, et plus particulièrement à se multiplier, à s'agréger ou à migrer, pendant les périodes de pullulation.

## Abstract

### Population genetics of an outbreaking insect, the migratory locust, *Locusta migratoria*

Some forest and crop pest insects are subject to pronounced and unpredictable fluctuations in population density. Studies of population evolution, focusing on neutral markers and density-dependent phenotypic traits, are recurrently cited as promising candidates to assess (i) the population dynamics and structure of outbreaking populations and (ii) the populational and environmental processes driving outbreak events, of these harmful species. These questions remain, however, largely unstudied in pest insects, in part because of the high prevalence of null alleles at microsatellite markers in these species. By using computer simulations, we measured here the population differentiation overestimation and the within-population genetic variation underestimation due to the presence of null alleles. We suggested an unbiased $F_{ST}$ estimator, restricted to visible allele states once genotype frequencies had been corrected. This work on null alleles enabled us to better describe the genetic variation at microsatellite markers within and among populations of the migratory locust, *Locusta migratoria*, that displays, at irregular intervals, huge increases of population densities with actively aggregating and swarming individuals typical of a high density distinctive form, the gregarious phase. The worldwide genetic structure among populations of this cosmopolitan species was found to be substantially incongruent with the current taxonomy in eleven subspecies based on morphometric criteria. Geographical and ecological barriers to gene flow as well as historical events are likely to drive patterns of genetic variation among populations at the worldwide scale. On the other hand, the neutral population structure did not show imprints of contrasted propensities to outbreak of *L. migratoria* populations. At a regional scale, our microsatellite survey provided strong evidence of a homogenizing effect of recent outbreak events at large spatial scales. We also carried out laboratory experiments to provide evolutionary and demographical insight into the phase change in historically outbreaking environments. We found non ambiguous evidence that a historically outbreaking Malagasy population of *L. migratoria* expressed larger behavioural and morphometrical phase changes than a historically non-outbreaking French population, presumably as a result of a genetically-based adaptive process. Moreover, the historically outbreaking Malagasy population invested more in first reproduction when gregarized, by both earlier reproduction and higher offspring quality, suggesting that gregarization may be a population regulating process that could explain the formation and/or duration of outbreaks in *L. migratoria*. In conclusion, our study prevents for the lack of resolution of molecular approaches for inferring the origin of nascent outbreak populations of *L. migratoria* and their subsequent movements due to low genetic differentiation between populations in historically outbreaking areas. However, it opens possibilities for engineering a genomics-based locust control, by silencing the expression of genes implied in the ability to gregarize, and in particular to grow, to aggregate, and to migrate, of outbreaking populations.

**Spécialité** : Biologie des populations et écologie
**Ecole Doctorale** : Biologie intégrative
**Formation Doctorale** : Biologie de l'évolution et écologie
**Mots clefs** : allèles nuls, changement de phase densité-dépendant, comportement, effets maternels, évolution expérimentale, génétique des populations, insecte, locuste, marqueurs microsatellites, phylogéographie, pullulation, traits d'histoire de vie
**Keywords** : behaviour, density-dependent phase change, experimental evolution, insect, life history traits, locust, maternal effects, microsatellite markers, null alleles, outbreak event, population genetics, phylogeography

Centre de Biologie et de Gestion des Populations, Campus International de Baillarguet, CS 30016, 34988 Montferrier / Lez
Génétique et Evolution des Maladies Infectieuses, UMR 2724, 911 avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5
CIRAD, UPR Ecologie et Maîtrise des Populations d'Acridiens, Campus International de Baillarguet, 34398 Montpellier Cedex 5